Decades after Voyager, a new bid to explore interstellar space p. 460

Adhesives for wearable ultrasound pp. 466 & 517

Pathogen detection linked to immune response pp. 486–488

Science

$15
29 JULY 2022
science.org

AAAS

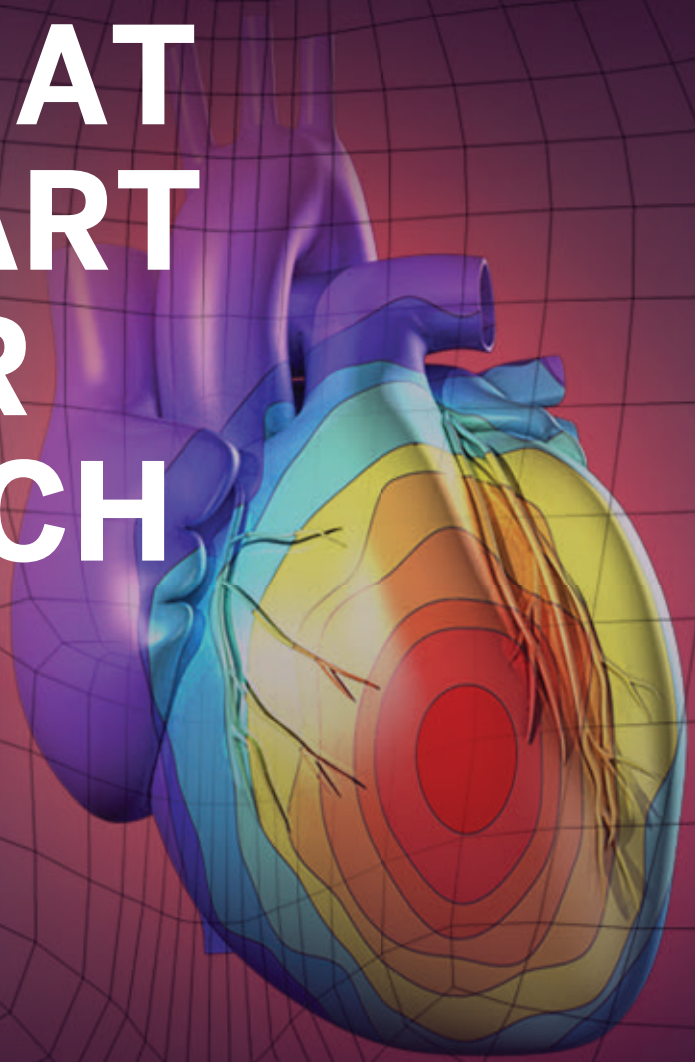
POLLINATORS OF THE SEA

Isopods fertilize red algae

pp. 471 & 528

science.org/journal/stm

PUT HUMAN HEALTH AT THE HEART OF YOUR RESEARCH



Submit your research:
cts.ScienceMag.org

Science Translational Medicine
AAAS

Twitter: @ScienceTM
Facebook: @ScienceTranslationalMedicine

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

CONTENTS

29 JULY 2022 • VOLUME 377 • ISSUE 6605

460

NEWS

IN BRIEF

452 News at a glance

IN DEPTH

454 Surprise virus tied to pediatric hepatitis cases
Two viruses plus a child's genetic background may explain a recent surge in the United Kingdom *By J. Kaiser*

455 NSF grant decisions reflect systemic racism, study argues
Success rates for white scientists far exceed the NSF average, whereas Black and Asian researchers do worse
By J. Mervis

456 Ancient Europeans farmed dairy—but couldn't digest milk
Giant study of ancient pottery and DNA challenges common evolutionary explanation for lactase persistence
By C. O'Grady

457 Radio bursts from 'zombie' black holes excite astronomers
Delayed emissions from black holes that fed on stars earlier could help explain the formation of powerful jets
By Z. Savitsky

458 The most unusual birds are also the most at risk
Two studies predict homogenization of the avian world as climate change and other human impacts continue
By E. Pennisi

459 Antibiotic after sex could help curb infections
Doxycycline reduces risk of sexually transmitted diseases—but will it fuel drug resistance? *By J. Cohen*

FEATURES

460 The long shot
The Voyager probes raised perplexing questions as they exited the Solar System. Now, scientists have conceived new missions to interstellar space *By R. Stone*
PODCAST

INSIGHTS

PERSPECTIVES

466 Seeing inside a body in motion
Adhesive ultrasound patches can provide medical imaging for patients on the go
By P. Tan and N. Lu
REPORT p. 517

467 Targeting brain cancer
A new drug is designed to hit an old target with increased precision *By R. Reddel and A. Aref*
RESEARCH ARTICLE p. 502

469 Molecular recorders to track cellular events
DNA tapes could be used to record dynamic molecular and cellular events in animals
By N. Masuyama et al.

471 Did pollination exist before plants?
Resarch shows that seaweeds depend on crustaceans for fertilization
By J. Ollerton and Z.-X. Ren
REPORT p. 528

472 The material properties of chromatin in vivo
Chromatin is fluidlike within the crowded nucleus when probed in a living cell
By W. Y. So and K. Tanner
RESEARCH ARTICLE p. 489

473 Engineering defects with DNA
Single-walled carbon nanotubes are structurally modified by using a genetic sequence *By Y. Wang*
REPORT p. 535

POLICY FORUM

475 A treaty to break the pandemic cycle
An evidence-based treaty must balance prevention, preparedness, response, and repair *By A. L. Phelan and C. J. Carlson*

466 & 517

CREDITS: (ILLUSTRATION) JOHNS HOPKINS APL/MIKE YAKOVLEV; (PHOTO) C. WANG *ET AL.*

BOOKS *ET AL.*

478 The hidden costs of batteries
Clean energy plans must include safer, more sustainable, worker-friendly power storage
By B. K. Sovacool

479 The ascent of Moderna
From scrappy startup to household name, a journalist traces the trajectory of the company and its mRNA vaccine *By E. Dolgin*

LETTERS

480 Funding databases for Ukrainian academics
By V. Mosienko et al.

480 Chinchilla conservation vs. gold mining in Chile
By J. E. Jiménez et al.

481 Life in science: A scuffle at sea
By J. Hernández-Urcera and M. E. Garci

RESEARCH

IN BRIEF

483 From *Science* and other journals

REVIEW

486 Innate immunity
Shared TIR enzymatic functions regulate cell death and immunity across the tree of life
K. Essuman et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABO0001
RESEARCH ARTICLES pp. 487 & 488

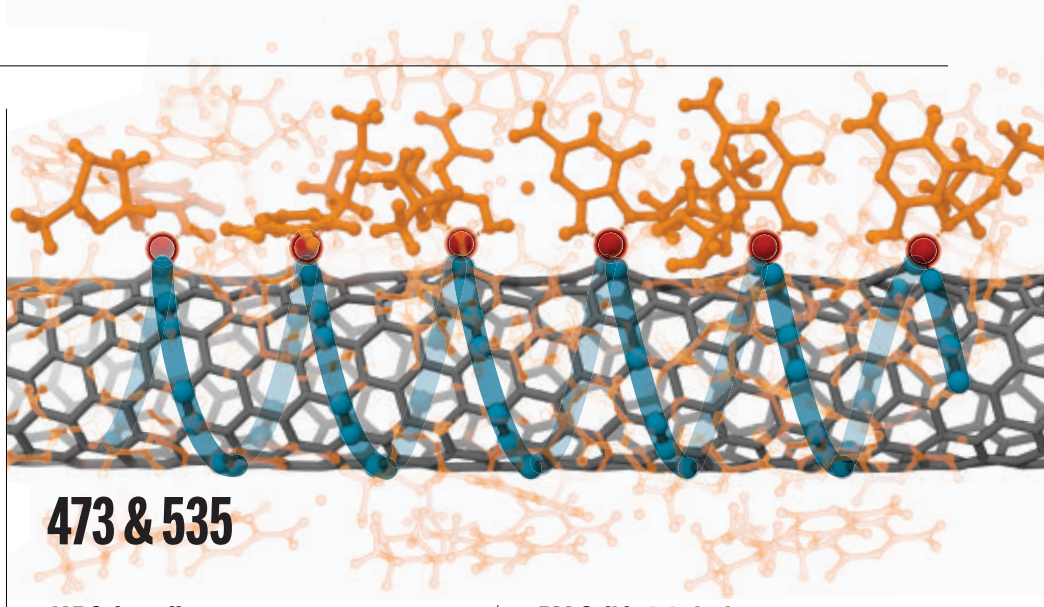
RESEARCH ARTICLES

Plant science

487 Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity *S. Huang* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ3297

488 TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity *A. Jia* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ8180
REVIEW p. 486

489 Biophysics
Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics *V. I. P. Keizer* et al.
PERSPECTIVE p. 472



473 & 535

495 Solar cells
Ion-modulated radical doping of spiro-OMeTAD for more efficient and stable perovskite solar cells *T. Zhang* et al.

502 Cancer
Mechanism-based design of agents that selectively target drug-resistant glioma
K. Lin et al.
PERSPECTIVE p. 467

511 Neurodevelopment
Analysis of somatic mutations in 131 human brains reveals aging-associated hypermutability *T. Bae* et al.

REPORTS

517 Bioimaging
Bioadhesive ultrasound for long-term continuous imaging of diverse organs
C. Wang et al.
PERSPECTIVE p. 466

523 Coastal ecology
Constraints on the adjustment of tidal marshes to accelerating sea level rise
N. Saintilan et al.

528 Species interactions
Pollinators of the sea: A discovery of animal-mediated fertilization in seaweed
E. Lavaut et al.
PERSPECTIVE p. 471; PODCAST

531 Solar cells
Inactive $(PbI_2)_2RbCl$ stabilizes perovskite films for efficient solar cells *Y. Zhao* et al.

535 Carbon nanotubes
DNA-guided lattice remodeling of carbon nanotubes *Z. Lin* et al.
PERSPECTIVE p. 473

539 Solid-state ionics
Nanosecond protonic programmable resistors for analog deep learning
M. Onen et al.

543 Cilia and flagella
In situ architecture of the ciliary base reveals the stepwise assembly of intraflagellar transport trains *H. van den Hoek* et al.

DEPARTMENTS

450 Editorial
How will EPA regulate the power sector?
By R. L. Revesz

451 Editorial
The court is lost
By H. H. Thorp

550 Working Life
Unequal undertakings
By A. Elbassiouny

ON THE COVER

An isopod (*Idotea balthica*), a type of crustacean, covers itself with male gametes while feeding on the epiphytes of the red alga *Gracilaria gracilis*. Isopods facilitate algal fertilization by carrying spermatia on their bodies. Such animal-mediated fertilization is akin to pollination in flowering plants. See pages 471 and 528. *Photo: Wilfried Thomas, Station Biologique de Roscoff*

AAAS News & Notes 482
Science Careers 549

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Copyright © 2022 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS.

Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

EDITORIAL

How will EPA regulate the power sector?

Richard L. Revesz is the AnBryce Professor of Law, Dean Emeritus, and director of the Institute for Policy Integrity at the New York University School of Law, New York, NY, USA. revesz@exchange.law.nyu.edu

As threats from climate change become more urgent, the US Supreme Court has responded by erecting a new roadblock to effective climate policy. Last month, it struck down the Clean Power Plan, the Obama administration's never-implemented regulation of greenhouse gas emissions from existing power plants. The ruling [*West Virginia v. Environmental Protection Agency* (EPA)] is a blow to climate action and could signal the court's hostility to a wide range of future regulations within and beyond the climate and environmental sphere, including those related to consumer protection and worker safety. Although the immediate effects on US climate policy aren't pervasive, EPA now needs to evaluate the emissions-reduction potential and legal risks of alternative regulatory approaches for the power sector.

The court's decision didn't question EPA's authority to regulate greenhouse gas emissions under the Clean Air Act. Similarly, it didn't limit authority to regulate greenhouse gas emissions from existing power plants under section 111(d) of the Clean Air Act—the provision used in the Clean Power Plan. The direct impact of the decision is relatively narrow, foreclosing only one approach to power sector climate regulation. This leaves EPA with other options.

> "The ruling... is a blow to climate action..."

The court objected to the Clean Power Plan's reliance on "generation shifting," a partial shift in electricity production from coal to natural gas, which generates fewer emissions, and from coal and natural gas to renewables, which produce no emissions. According to the court, the Clean Air Act didn't authorize EPA to base its standard on this technique. Climate regulations for other sectors, including transportation, oil and gas, and manufacturing, are likely to be source specific, and therefore unaffected by this decision.

With generation shifting now off the table, EPA must choose the "best system of emission reduction" under the relevant provision in the Clean Air Act. This system must be "adequately demonstrated," which means that it cannot be too speculative. And the agency must "take into account the cost of achieving such reduction," which means that costs cannot be excessive. EPA could consider basing the "system" on technologies such as carbon capture and storage (CCS). It could also consider requiring the combustion of coal simultaneously with natural gas at coal plants (cofiring), which reduces emissions by replacing some coal with natural gas.

CCS, which captures and stores carbon dioxide before it is released into the atmosphere, will lead to the greatest emission reductions among the options remaining. Thus, EPA is compelled to adopt it as long as other statutory requirements are met. In 2017, a document filed in a federal litigation on greenhouse gas standards strongly argued that the technology was "adequately demonstrated," and there have since been further technological developments. The process is costly but the 2021 infrastructure bill provides substantial subsidies for its use, and more federal money is possible through future congressional action. EPA will need to evaluate whether the funding sufficiently lowers the effective cost of CCS. Because some states have higher subsidies for this technology than others, EPA could divide the power sector into regions and require CCS in states with larger subsidies.

Should EPA conclude that the costs of universal CCS are excessive, it could require power plants with multiple contiguous generating units to partially adopt CCS. For example, EPA standards could require plants with three or more contiguous units to use CCS at one-third of the units. This technique, called "bubbling," is a common regulatory tool upheld by the Supreme Court. EPA could also evaluate whether the statute permits this approach for noncontiguous sources.

As an alternative, the agency could require that coal plants cofire with natural gas. Because many coal plants have already adopted cofiring, it is clearly "adequately demonstrated" and doesn't require that "coal plants...become natural gas plants"—which the Supreme Court indicated would be problematic. If EPA were concerned about the legality of an across-the-board cofiring standard, the agency could restrict this requirement to a subcategory of the industry that already uses cofiring and set standards for other sources using a different approach.

Generation shifting would have reduced emissions at the lowest cost through a cap-and-trade program or related regime, but that's no longer possible. Unfortunately, other regulatory approaches for the power sector will be more costly or less effective. Moreover, these regulations, as well as other regulations in the climate change and environmental sectors, may be slowed or stalled on other grounds. The fallout from the Supreme Court's decision across the economy is likely to have only just begun.

–Richard L. Revesz

10.1126/science.ade0779

PHOTO: NEW YORK UNIVERSITY SCHOOL OF LAW

The court is lost

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

The United States has an insatiable desire for technological advancement but is governed by founding documents that are completely unsuited for science and technology. This incongruity has manifested in recent disastrous actions by the US Supreme Court on guns, abortion, and climate. The decisions suggest that the battle is being won by the portion of America who—while lionizing the past and clinging to the infallibility of words written in the late 18th century—can't put down their cell phones. Reactionary posts on social media wouldn't get very far without a hundred years of technical advances—and massive amounts of power to recharge mobile device batteries and run the server farms that support the digital world. Because the disconnect between aspects of modern life and the framing of the country's governance appears inconsequential to the conservative majority of justices in the US Supreme Court, it is vital that the scientific community advocate a political and societal landscape in which compassion and adaptability attend technological progress.

Some of the most compelling writing on this topic has come from Harvard historian Jill Lepore. Lepore wrote a prescient *Science* editorial before the 2020 US presidential election in which she made the case that accurately counting the votes in the election was more important than a quick tally. In her sweeping and compelling history of the United States, *These Truths*, Lepore chronicles the political impacts of America's obsession with technological progress that began with the industrial revolution and continued through space exploration, nuclear weapons, and the internet. Science and technology marched ahead against a backdrop of devotion to a set of founding documents that didn't contemplate any of these advances.

"The Constitution establishes fundamental law, the law that governs the government," Lepore told me. "At the time it was written, in 1787, women couldn't run for office, or vote; they were shunned even for expressing political views. At the first federal census in 1790, nearly one in six people were held as property. Bearing 'arms' at the time the Bill of Rights was ratified, in 1791, meant, mostly, hauling around a lousy old musket that took ten minutes to load in hopes of shooting a rabbit for supper."

"The court's recent spate of bad decisions mocks scientific facts."

The court's recent spate of bad decisions mocks scientific facts. It obviated a modest and reasonable gun control law in New York despite research showing that gun control saves lives. It removed a woman's right to abortion even though the Turnaway Study and others like it show that access to abortion improves health. They attributed their rulings to fealty to a document written by slaveowners who had no idea about automatic weapons and who didn't even think women should participate in the government.

The recent decision on the Environmental Protection Agency (EPA) is a body blow for environmental science. It struck down the implementation of a completely reasonable plan that allowed the agency to regulate greenhouse gas emissions from existing power plants. The court objected to the fact that the plan would have shifted production from coal to natural gas. Apparently, they couldn't find any support in the Constitution for environmental regulation. The notion that the Constitution would contemplate climate change is ridiculous. As Lepore pointed out, "The average annual temperature in Philadelphia was fifty-two degrees Fahrenheit. By 1962, the year Rachel Carson published *Silent Spring*, it was fifty-four; by 2017, the year the Trump Administration withdrew the United States from the Paris Climate Accords, it was fifty-nine."

The scientific enterprise has contributed to the tension between the conservative court and a developed society by enjoying the psychic and monetary rewards for its technological achievements without demanding social responsibility in return. Science must realize that progress is more than just filling journals with knowledge and hoping for the best. The scientific community must value and partner with communicators and policymakers who can help show that scientific advancement demands that the nation operate as a work in progress. Otherwise, America will be stuck with a government that worships a set of documents created by men who had no idea about evolution, dinosaurs, hydrocarbons, women's health, or digital communication. "There's a great deal to be said for constitutional stability, for relying on the constitution as the expression of a set of foundational principles," said Lepore. "But reading it as a Bible for solving the problems of the 21st century? The shortcomings are overwhelming."

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.ade0780

NEWS

IN BRIEF

Edited by
Jeffrey Brainard

New telescopes will rapidly scan the sky to facilitate timely observations of areas where gravitational waves originated.

ASTRONOMY

Speedy scopes to spy gravitational wave sources

Researchers last week reached the midpoint in building a pair of observatories designed to pinpoint the location of cataclysmic events sensed by gravitational wave detectors so that other astronomers can quickly zoom in on the aftermath. The Gravitational-wave Optical Transient Observer (GOTO) uses two sets of 16 small telescopes, one in the Canary Islands—now operational—and one in Australia, whose construction has just started. They will swing into action automatically when gravitational wave detectors, such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) in the United States and Virgo in Italy, register a space-time ripple caused by events such as a merger of two black holes. Such detectors don't give precise locations, so GOTO's scopes will sweep the region of space for rapidly brightening objects that could be the source of the wave; operators will then send alerts, giving more sensitive optical telescopes a location to aim at. A neutron star merger, spotted gravitationally in 2017, produced visible signals, but telescopes took 11 hours to find them after the event became known, and missed the explosion's early phases. The £4.4 million GOTO telescopes hope to do better and flag candidates in half an hour. Project leaders aim to be ready by March 2023, when upgrades to LIGO and Virgo are finished and they begin their next observing runs.

Monkeypox emergency declared

INFECTIOUS DISEASES | The World Health Organization (WHO) last week declared the global monkeypox outbreak, which has sickened more than 15,000 in at least 70 countries, a Public Health Emergency of International Concern (PHEIC), even though a sharply divided advisory committee had not recommended doing so. In June, WHO's Emergency Committee for monkeypox first advised against giving the epidemic PHEIC status—which grants WHO extra powers and helps focus political attention on an outbreak—a decision that was widely criticized. At its second meeting on 21 July, the panel could not reach a consensus, with nine members opposing a PHEIC declaration and six supporting it. (The meeting was followed by tense exchanges between participants via email and text messages, *Science* has learned.) Opponents noted that monkeypox is not yet circulating in the general population; it has been observed mostly in men who have sex with men. But others argued for action now because there is a danger the virus could become established long-term in the wider population. On 23 July, WHO Director-General Tedros Adhanom Ghebreyesus said the outbreak met criteria in the International Health Regulations and, in an unprecedented move, declared a PHEIC anyway.

NASA delays Moon rover launch

LUNAR SCIENCE | The first NASA robotic rover to visit the Moon will be delayed by a year, launching in November 2024, the agency announced last week. The Volatiles Investigating Polar Exploration Rover, known as VIPER, will land at the Moon's south pole, hunting for water ice. The delay, which will nudge the $754 million mission closer to the planned launch by China of a similar device, will allow further testing of the lunar lander, developed by Astrobotic Technology, that will carry the rover in the lander's first lunar touchdown.

'Plan B' for U.K. science funding

BREXIT | The U.K. government last week outlined a domestic program to replace funding that U.K.-based researchers would have had access to from the European Union's flagship research scheme. The EU has avoided finalizing a deal to allow U.K. participation in the 7-year, €95 billion

PHOTO: UNIVERSITY OF WARWICK

Horizon Europe program because of ongoing disputes over trade in Northern Ireland. The United Kingdom's "Plan B" would kick in if the Horizon deal completely falls through. U.K. researchers could continue to participate in Horizon-funded international collaborations that include at least three partners in EU member states or other Horizon-associated countries. EU rules would require these U.K. researchers to bring their own funding to the table, and Plan B would put no immediate limits on funding such consortium participation.

Trump policy on species reversed

CONSERVATION | Completing a review President Joe Biden promised on his first day in office, the U.S. Fish and Wildlife Service (FWS) last week finished rolling back changes made to the Endangered Species Act by former President Donald Trump's administration. One of Trump's changes weakened a provision that governs when property owners may damage critical habitat needed by endangered species to survive. The Trump rule would have required FWS to allow landowners to do such damage—by filling wetlands, for example—if they showed credible evidence that continuing to protect the critical habitat would cause economic harm. On 21 July, the Biden administration reinstated a previous rule that allowed FWS to use its discretion on whether to grant such requests. In June, the agency undid a separate Trump-era decision that raised the bar for designating land as critical habitat.

Surgeon to head cancer institute

LEADERSHIP | President Joe Biden is expected to pick cancer surgeon Monica Bertagnolli as the next director of the National Cancer Institute (NCI), the world's largest funder of cancer research. Bertagnolli, a physician-scientist at Brigham and Women's Hospital, the Dana-Farber Brigham Cancer Center, and Harvard Medical School, specializes in gastrointestinal cancers and is well known for her expertise in clinical trials. Her lab research on tumor immunology and the role of a gene called *APC* in colorectal cancer led to a landmark trial she headed showing an anti-inflammatory drug can help prevent this cancer. Bertagnolli, 63, will be the first woman to lead NCI.

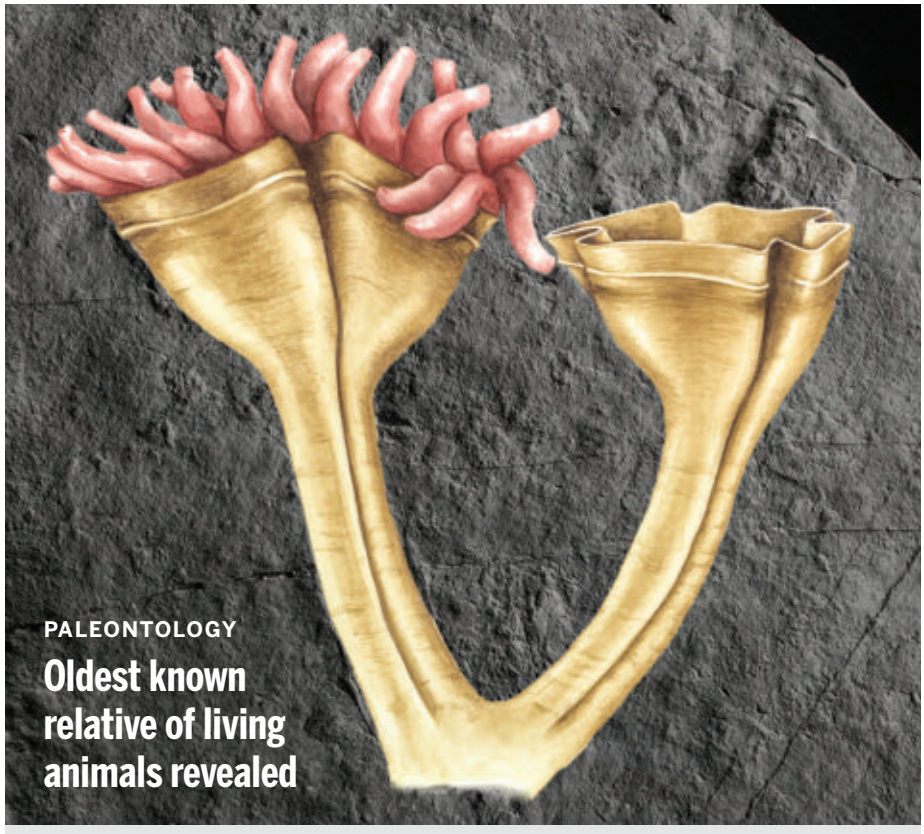
A head for U.K. 'high-risk' funder

LEADERSHIP | Ilan Gur, an entrepreneur and materials scientist, will lead the United Kingdom's new agency to fund outside-the-box research leading to commercial spinoffs. The government this year created the Advanced Research and Invention Agency (ARIA), modeled on the U.S. Defense Advanced Research Projects Agency (DARPA) and its sponsorship of "high-risk, high-reward" science. In February, ARIA selected DARPA's deputy director, Peter Highnam, as its head, but he withdrew before starting the job. Gur worked as a program officer at ARPA-E, the U.S. Department of Energy's version of DARPA. He also founded the U.S. nonprofit Activate, which has supported the creation of more than 100 science-based startups, according to a government statement. To encourage innovative grantmaking, ARIA's £800 million, 4-year budget will be separate from that of the country's main research funder, UK Research and Innovation; the sum amounts to only 1% of total R&D spending. Unlike DARPA and ARPA-E, ARIA so far lacks a specific research focus, which some observers have predicted will undermine its effectiveness.

THEY SAID IT

> "[He] was undoubtedly the best fossil hunter the world has ever had."

The National Museums of Kenya, in *The Star*, on the death of Kamoya Kimeu, 81, a Kenyan who found many important hominin fossils including "Turkana Boy," an almost complete *Homo erectus* skeleton, in 1984.



PALEONTOLOGY

Oldest known relative of living animals revealed

A fossil of an organism between 556 million and 562 million years old bears features resembling those of modern jellyfish and coral, and appears to be the oldest example of an evolutionary group still living today. Paleobiologist Philip Wilby and colleagues at the British Geological Survey discovered the new fossil in the Charnwood Forest, a hotbed of pre-Cambrian paleontology in central England. The organism predates the Cambrian explosion of about 539 million years ago, which scientists historically accepted as the origin of modern animals. The specimen, *Auroralumina attenboroughii* (artist's reconstruction, above), belongs to a group called cnidarians, which includes today's jellyfish and corals, the research team concludes this week in *Nature Ecology & Evolution*. The symmetry of its body is like that of modern jellyfish. The team selected its name in part to honor naturalist and broadcaster David Attenborough, who spent his childhood near Charnwood and has brought attention to its fossils.

IMAGE: BRITISH GEOLOGICAL SURVEY/© UK RESEARCH AND INNOVATION

In March, 3-year-old Lola-Rose Raine of Gravesend, England, received a liver transplant from her father after she developed severe, unexplained hepatitis.

PHOTO: ALAN RAINE

IN DEPTH

HEALTH

Surprise virus tied to pediatric hepatitis cases

Two viruses plus a child's genetic background may explain a recent surge in the United Kingdom

By **Jocelyn Kaiser**

Starting in April, young children with mysterious cases of acute liver inflammation began to attract scientific—and media—attention in the United Kingdom and elsewhere. Most would recover, some after liver transplants, but a few died. Physicians could find no evidence in the children of viruses that typically cause hepatitis, but researchers homed in on an unexpected suspect: adenovirus, a family of cold-causing viruses common in kids. Now, a sweeping genetic search of these patients instead implicates another virus, one previously thought to be completely harmless.

In two independent, preliminary studies, U.K. researchers found high adeno-associated virus 2 (AAV2) levels in the blood or liver cells in 24 of 25 children with unexplained hepatitis. In a group of kids without this condition, almost none had AAV2, even those with adenovirus. The young hepatitis patients were also much likelier to have a genetic variant that may make their immune systems overreact to viruses.

The preprints, one posted on medRxiv on 19 July and the other just submitted there, offer new clues to the mystery pediatric hepatitis, which now numbers more than 1000 probable cases in 35 countries and has caused at least 22 deaths.

> **"It's intriguing for sure, but it's a very small number of cases and controls."**
>
> **Saul Karpen**,
> Emory University School of Medicine

But outside researchers are cautious about the studies, which haven't been peer reviewed. "It's intriguing for sure, but it's a very small number of cases and controls. It could be associations and not causation," says Saul Karpen, a pediatric gastroenterologist at Emory University School of Medicine.

Severe hepatitis in children is rare and often goes unexplained. But alarm bells went off in the spring after hospitals in the United Kingdom reported a rise in cases—now up to 272—that are not linked to the usual causes, such as a hepatitis A, B, or C virus. Another 334 such unexplained cases have been reported in the United States, but the Centers for Disease Control and Prevention said in June that these are not above usual levels.

Still, a spring spike in adenovirus infections in the United Kingdom lent credibility to a hepatitis surge there. And last week in two papers in *The New England Journal of Medicine*, teams in Birmingham, England, and Birmingham, Alabama, reported adenovirus 41, a specific type that causes gastrointestinal illness, in the majority of nine U.S. and 30 U.K. cases tested.

But the London and Glasgow, Scotland, teams behind the two preprints wanted to make sure some new virus or viral variant wasn't emerging. They used DNA and RNA sequencing of tissue samples from U.K. cases to look for about 200 different virus families. AAV2 jumped out. It was in liver or blood samples in all nine Scottish cases examined and 15 of 16 separate cases from all over the United Kingdom, according to the studies. In contrast, almost none of 158 controls, which included healthy children and children with hepatitis for known reasons, had AAV2.

The viral family to which AAV2 belongs was initially discovered in samples that contained adenoviruses—AAVs need a protein from a distant viral relative, usually adenovirus or herpesvirus, to replicate. Many people are infected by AAV2 by age 10, but the virus can lie dormant in cells until that helper virus comes along and activates it.

Most of the patients also had adenovirus, usually the 41 type. But AAV2 never turned up in a control group of 12 Scottish children who had an adenovirus infection but no hepatitis. The authors of the preprints also ruled out SARS-CoV-2, an early suspect, partly because its infection

rates weren't higher than average in the Scottish hepatitis patients.

The pandemic could, however, indirectly explain the rise of pediatric hepatitis in the United Kingdom. Relaxed COVID-19 restrictions meant children may have been exposed to a "cocktail of viruses" all at once versus gradually, according to Emma Thomson, an infectious diseases specialist at the MRC-University of Glasgow Centre for Virus Research and a leader of the Scottish study.

Both studies also found that all but two of 14 sick children tested carried a specific version of a type of gene called *HLA* that helps shape the body's response to pathogens. The variant is particularly common in northern Europeans—16% of Scottish people carry it, and it is known to be linked to some autoimmune disorders, Thomson notes.

Although the researchers found genetic material from AAV2 in patient's liver cells, they did not detect viral proteins or actual copies of the virus. That suggests instead of directly damaging liver cells, AAV2 may provoke an immune response that harms the organ. That theory is supported by the link with *HLA* type and that some kids had gastrointestinal illnesses many weeks before they developed hepatitis, said a leader of the London team, virologist Judy Breuer of Great Ormond Street Hospital and the University College London Great Ormond Street Institute of Child Health, at a press briefing Monday.

A pathogenic role for AAV2 "goes against everything we know" about the virus, says virologist Eric Kremer of the Institute of Molecular Genetics of Montpellier. Looking at the type of antibodies to AAV2 in the children, he notes, could support its guilt by distinguishing between a long-ago infection, or a more recent one timed with the onset of hepatitis. Those studies are planned, Thomson says.

The good news for U.K. parents in all this: Pediatric hepatitis cases have dropped in recent weeks to "almost ... background," said Meera Chand of the UK Health Security Agency, which expects to release its own study of adenoviruses in these cases this week.

At the same time, the studies raise questions about delivering therapeutic genes with AAVs, as many groups are trying to, Kremer notes. Although the viruses are further modified so they can't replicate even if a helper virus is around, they have sometimes caused liver inflammation—which in rare cases may have contributed to deaths. Stanford University gene therapy researcher Mark Kay says the field may want to explore whether *HLA* type can predict whether an AAV gene therapy patient will experience liver toxicity. ■

DATA: C. CHEN *ET AL.*, OSF PREPRINTS (2022) DOI:10.31219/OSF.IO/XB57U

SCIENTIFIC COMMUNITY

NSF grant decisions reflect systemic racism, study argues

Success rates for white scientists far exceed the NSF average, whereas Black and Asian researchers do worse

By **Jeffrey Mervis**

White scientists are more likely to win a grant from the National Science Foundation than researchers from other racial and ethnic groups, according to an independent analysis of more than 2 decades of NSF data on its merit review process.

The analysis supports earlier studies finding similar racial disparities in the funding of scientists by other federal agencies, notably the National Institutes of Health (NIH). And its authors—a team led by geochemist Christine Yifeng Chen, a postdoc at Lawrence Livermore National Laboratory—attribute the gap in NSF funding rates, with white scientists at the top and Asian researchers at the bottom, to "systemic racism."

The NSF funding disparities "have cascading impacts that perpetuate a cumulative advantage to White [principal investigators] across all of science, technology, engineering, and mathematics," they write in their study, posted earlier this month on the Center for Open Science preprint site.

The team gave a copy of its analysis to NSF leadership, which is not challenging its conclusions. NSF Director Sethuraman Panchanathan "shares these concerns [about] systemic racial disparities in funding at NSF and other federal agencies," an agency spokesperson says.

The researchers delved into detailed annual NSF reports that contain data on 1 million proposals submitted to the agency between 1996 and 2019. The reports mention racial disparities only in passing, but Chen and her team focused on them after hearing complaints from senior nonwhite colleagues about what they felt was an uneven playing field in winning NSF grants.

"I think it's significant that this project was initiated by early-career scientists," Chen says. "It speaks to the prevailing culture in academia that allows the status quo to be perpetuated. We felt that if we didn't do the analysis, nobody else would."

Every scientist submitting a proposal to NSF faces stiff competition; overall success rates fluctuated between 22% and 34% over the study period. But white scientists consistently did better, Chen and colleagues found. (Scientists are asked to voluntarily provide their race and ethnicity when applying for a grant.)

In 2019, for example, NSF funded 31.3% of proposals from white scientists, versus an overall rate of 27.4%. In contrast, the success rate was 22.4% for Asian scientists and 26.5% for Black scientists. Proposals from Latino scientists were funded 29% of the time, a rate slightly above average but below the rate for white scientists.

Chen and colleagues translate the higher success rates into what they call "surplus awards." In 2019, when NSF received about 42,000 proposals, the team calculated that white scientists received 798 surplus grants. The cumulative surplus over 20 years was 12,820 awards (see table, left).

In contrast, Asian scientists received 460 fewer awards in 2019 than they would have had their success rate been comparable, with a cumulative "unfunded" total of 9701 awards. (Asian scientists submit the second most proposals of the various groups every year, roughly half the number sent in by white scientists.) The relative advantage for white scientists has steadily grown, the analysis shows, from 3 percentage points above the agency's average success rate in 1999 to 14 percentage points above in 2019.

For Black scientists, the funding gap over that period was smaller, but still significant.

Funding issue

A study of racial disparities in National Science Foundation funding concludes that white researchers reaped a large cumulative "surplus" of awards between 1999 and 2019, whereas Asian applicants experienced significant "underfunding," with other groups falling in between.

RACE/ETHNICITY	AWARDS
White	+12,820
American Indian/ Native Alaskan	+80
Native Hawaiian/ Pacific Islander	–17
Hispanic/Latino	–175
Black/African American	–417
Asian	–9701

(Black researchers submit one proposal for every 20 proposals NSF receives from white scientists.) The average funding rate for Black scientists was 8 percentage points below that of white scientists, according to the preprint's authors. Pacific Islander and Native Hawaiian scientists, an even smaller cohort of applicants, had success rates 11 percentage points below that of white scientists. In contrast, Latino scientists—who submit about 50% more proposals to NSF than do Black scientists—have done slightly better than the norm but were 2 percentage points below the success rate for white scientists in 2019.

About 75% of the proposals NSF receives each year are classified as research proposals. The remaining requests are to support education and training, equipment and facilities, conferences, and other activities. White scientists enjoy an even larger advantage over most other groups in winning research awards, the study found. For nonresearch awards, most nonwhite groups did better than the NSF-wide average.

In 2011, a team led by University of Kansas, Lawrence, economist Donna Ginther found similar racial disparities among NIH grant recipients, including a gap of up to 13 percentage points in success rates between white and Black scientists. That gap shrunk—but did not disappear—in follow-up studies that accounted for such factors as an applicant's publication history, prior funding, age, academic rank, and how much research takes place at their institution.

Even so, former NIH Director Francis Collins last year apologized to "individuals in the biomedical research enterprise who have endured disadvantages due to structural racism." NIH has tried to reduce disparities with programs designed to increase the success of Black applicants.

Unlike NIH with Ginther, NSF did not give Chen's team access to applicant data that would have allowed it to do such a multivariate analysis. "That information would have been extremely valuable for looking at issues of intersectionality," says co-author Aradhna Tripati, a geoscientist at the University of California, Los Angeles. "We also could have looked at the impact of NSF's existing programs to foster equity and broader participation in science."

Even without that extra level of analysis, the study's findings appear sound, says Susan White, director of the statistical research center at the American Institute of Physics. "It may not be exactly the number [the authors] present, but I don't doubt that the disparity is real, and serious," she says.

An NSF spokesperson says although the agency is proud of its array of programs designed to address equity and inclusion, "there is still much [work] to do." ■

HUMAN EVOLUTION

Ancient Europeans farmed dairy—but couldn't digest milk

Giant study of ancient pottery and DNA challenges common evolutionary explanation for lactase persistence

By **Cathleen O'Grady**

Over the past 10,000 years, populations living far apart in Europe, Africa, South Asia, and the Middle East separately acquired a key genetic change: the ability to digest the milk sugar lactose as adults. Researchers thought people who had that ability and lived in dairy farming cultures got a nutritional boost and had more children, thus spreading the genetic changes.

But in recent years, unexpected findings—such as data from Mongolia, where people devour milk products but 95% of adults are genetically lactose intolerant—have challenged that story. Now, a study combines large archaeological data sets on dairy farming with ancient DNA and finds that across Europe, people consumed dairy for millennia before lactase persistence into adulthood was widespread. The researchers suggest illness and famine may have turned lactose intolerance from uncomfortable to deadly, driving periods of intense selection for the digestive trait.

A farmer in Mongolia, where most people are lactose intolerant, milks a cow.

PHOTO: TAYLOR WEIDMAN/BLOOMBERG/GETTY IMAGES

The study "changes our long-term understanding of the relationship between milk use and lactase persistence," says Jessica Hendy, an archaeologist at the University of York who was not involved in the work.

In the new study, archaeologists compiled evidence of milk from nearly 7000 pieces of ancient pottery, taken from 554 European sites representing the past 9000 years. They tracked the rise and fall of dairy farming across Europe by analyzing the fats preserved in the pottery. With ancient DNA specialists, they then compared this with signs of lactase persistence in 1293 published human genomes from the same regions and period.

Fluctuating dairy use over time didn't match up with changes in lactase persistence. Instead, the researchers found that what they considered signals of famine and sickness best matched the jumps in lactase persistence in ancient DNA, they report in *Nature* this week. (They used archaeological records to identify periods of shrinking populations—perhaps famines—as well as times of greater population density—possibly times of faster disease spread.)

Lactose intolerance in dairying cultures might be dangerous for people who were sick or starving, suggests co-author Mark Thomas, a human evolutionary geneticist at University College London. A lactose intolerant person consuming milk normally suffers flatulence and diarrhea, with no more severe effects than embarrassment and discomfort, Thomas says: "But if you get diarrhea when you're severely malnourished, then you have serious problems. One of the biggest causes of death in the world is fluid loss in severely malnourished people."

The findings support the idea that dairy farming alone wasn't the key force behind the spread of lactase persistence, the researchers say: The selection pressure likely only grew strong when combined with sickness and starvation.

It's an "exciting avenue" for ongoing research, Hendy says. But she cautions it's difficult to estimate ancient population fluctuations and understand what led to them.

The research complements previous results, such as the puzzlingly late arrival of widespread lactase persistence in Central Europe, says Christina Warinner, a molecular archaeologist at Harvard University. But she says the new study brings the heft of several large data sets to the question. The story of dairy farming has been "full of surprises," Warinner says. "It's helping us to really appreciate better the complexities of the past." ■

Supermassive black holes can produce powerful jets when shredding nearby stars, as seen in this artist's conception. Now, astronomers are seeing secondary bursts months or years after the fact.

IMAGE: SOPHIA DAGNELLO; NRAO/AUI/NSF

ASTRONOMY

Radio bursts from 'zombie' black holes excite astronomers

Delayed emissions from black holes that fed on stars earlier could help explain the formation of powerful jets

By **Zack Savitsky**

When a hapless star ventures too close to one of the supermassive black holes that lurk at the center of galaxies, it's torn to shreds and stretched like spaghetti. In this so-called tidal disruption event (TDE), the black hole dines on the stellar remnants, which wrap around the black hole's belly in an accretion disk. During the feast, the black hole can glow brighter than a supernova for months, before returning to a quiet state of hibernation.

Or so the story usually goes.

Continued monitoring by patient astronomers has now revealed a few cases in which black holes wake up and belch matter and energy, sending bursts of radio waves toward Earth months or even years after the initial TDE. "What's incredibly unusual about [these events] is that the objects came back to life, like a zombie," says Enrico Ramirez-Ruiz, a theoretical astrophysicist at the University of California, Santa Cruz. "This is really challenging the paradigm."

Astronomers aren't sure what's triggering the delayed outbursts, but they think the emissions could help explain the mysterious mechanisms by which black holes convert infalling stellar material into powerful jets that rocket out from their poles. "It's telling us something about the physics of the central engine that's otherwise hidden from us," says Sasha Tchekhovskoy, a computational astrophysicist at Northwestern University. "These jets can explode entire galaxies, so it's a really important process in the evolution of galaxies."

Most of the few dozen known TDEs have been detected from the optical light or x-rays emitted in the initial feast. But, "Radio is now playing a very important role" in understanding TDEs, says astronomer Igor Andreoni of the Joint Space-Science Institute. Black holes generate radio waves by expelling plasma—pumping it out in polar jets or belching out material that crashes into surrounding gas. But these outflows normally take place during a TDE, shortly after the black hole rips apart its meal.

In February 2021, however, Assaf Horesh, an astrophysicist at the Hebrew University of Jerusalem, discovered a radio burst that came 6 months after the initial TDE. Then, on 30 June, Yvette Cendes, an astronomer at the Harvard & Smithsonian Center for Astrophysics, reported finding another delayed flare in a preprint posted to arXiv. Using multiple observatories, she and her colleagues documented a rapid spike in radio activity that launched more than 2 years after the black hole's initial snack. "It's a pretty exceptional case," Cendes says.

Horesh's graduate student Itai Sfaradi may have caught a third example. Reanalyzing a previously spotted TDE, Sfaradi claims in the 10 July issue of *The Astrophysical Journal* that he found delayed radio emissions in combination with an x-ray flare. These tandem emissions are sometimes seen in so-called x-ray binaries—in which star-size black holes suck gas away from a paired star—hinting that the mechanisms may be related, Horesh says.

Shifts in the black hole's accretion disk power the flare-ups from x-ray binaries, and Ramirez-Ruiz thinks the same thing may be happening with the supermassive black holes, months after a TDE. In this scenario, a star's spaghettified gas piles up more slowly over time, causing the accretion disk to grow colder and thinner. Eventually, the disk weakens enough to open an escape path that allows the black hole's magnetic field lines to launch material from the disk into space, where it crashes into surrounding gas and produces radio bursts.

Tchekhovskoy agrees—and he's got models that demonstrate the behavior. He and colleagues ran computer simulations of accretion disk evolution and found they can reach a Goldilocks state in which jets can form efficiently. The key moment comes when the accretion disk is still dense enough to fuel jets, but not so dense as to reabsorb the generated radio waves.

Perhaps that's why we're seeing these delayed bursts, he says—"We're just waiting for the gas to have the right density."

More clues could come if wide-field radio surveys can capture other zombie awakenings. The Very Large Array, a complex of telescopes in New Mexico, is set to scan the skies for the third time next year, and the Australian Square Kilometre Array Pathfinder will launch a full-sky survey later this year. Both Cendes and Horesh plan to conduct follow-up radio surveys of TDEs using these observatories, among others. In unpublished work, Cendes has already found what she thinks are several more candidates.

Discovering a larger population of these TDEs with delayed outbursts would unlock a natural laboratory, enabling theorists to investigate black hole behavior under a wide range of conditions, Ramirez-Ruiz says. For physicists, he says, "Black hole gastronomy really provides a new playground." ■

Birds of distinctive form and specialized ecological niches, such as these Himalayan griffon vultures in China, may be more likely to go extinct.

PHOTO: JED WEINGARTEN/WILD WONDERS OF CHINA/MINDEN PICTURES

BIODIVERSITY

The most unusual birds are also the most at risk

Two studies predict homogenization of the avian world as climate change and other human impacts continue

By **Elizabeth Pennisi**

It's bad enough that Earth could be losing thousands of species each year. Now, two independent studies of birds have concluded the ones most likely to disappear are those that serve unique—and possibly irreplaceable—functions in their ecosystems. Consider the toucan: Its iconic beak lets it eat and disperse seeds and fruit too large for other birds in South American rainforests. Yet these striking creatures, as well as vultures, ibises, and others with distinctive physical traits, are likely to be the first to go extinct, homogenizing the avian world, according to one study. A second predicts communities will grow more alike as the climate changes and species flock to cooler regions.

"That's alarming because we know that diversity of sizes and shapes and behaviors is a signature of a healthy community," says Scott Edwards, an evolutionary biologist at Harvard University. "This is laying out the grim world we are going to be facing."

He and others hope the studies will spur conservationists to think more broadly about what biodiversity means. "The impacts of human actions can actually be worse than what we might think just based on species tallies alone," says Jedediah Brodie, an ecologist at the University of Montana, Missoula. "We haven't focused on what kinds of functions we might be losing," adds macroecologist Marta Jarzyna of Ohio State University.

Every ecosystem depends on diverse organisms to fill a variety of roles. Among birds, some eat and disperse seeds whereas others eat carrion, helping recycle remains. Special traits aid these tasks: Pointed beaks help vultures tear into flesh whereas long legs keep wading birds' bodies dry. "When communities are homogenized, they lose a lot of those ecological functions," Brodie says.

To assess whether birds with certain sets of traits are disappearing, Emma Hughes, a newly minted Ph.D. at the University of Sheffield, spent several years measuring the beak size and shape, lower limb and wing length, and body size of museum specimens of almost 8500 birds from around the world. Working with conservation biologist David Edwards and macroecologist Gavin Thomas, both also at Sheffield, she used statistical techniques to chart similarities and differences among the species based on these traits. Many birds, such as songbirds, clustered together in shape. Outliers included big albatrosses and tiny hummingbirds.

Next, the researchers began to pare down this chart by removing the birds most likely to go extinct, using the International Union for Conservation of Nature's Red List of Threatened Species, which ranks species according to their probability of disappearing. The most threatened species also turned out to be the most distinctive in body shape and ecosystem function, the group reported last week in *Current Biology*. As the researchers sequentially removed species from most to least threatened, toucans, ibises, and other distinctive species dropped out; similar birds—think finches and starlings—remained.

To identify regions where the trend might hit hardest, the Sheffield team analyzed birds in 14 major habitats, such as tropical grasslands. They found that species homogenization will affect 12 of the 14 and will be most extreme in flooded grasslands and tropical forests. The "most imperiled" regions included southern Vietnam, Cambodia, and the Himalayan foothills, as well as islands such as Hawaii, which has already lost all its distinctive honeycreepers, Hughes notes. "In some cases, there are no other organisms that can replace the unique ecological roles that these species play," Brodie says.

A separate study led by Alke Voskamp, a macroecologist at the Senckenberg Biodiversity and Climate Research Centre, identified another homogenizing force: climate-driven shifts in birds' ranges. Her European team mapped the current ranges of 9882 bird species compiled by BirdLife International. Then they applied projections from global climate models to predict where those species could find hospitable habitat through 2080, and analyzed how the altered distributions would change the makeup of bird communities. "[Moving] reshuffles the communities," Voskamp says.

As expected, tropical and subtropical regions will likely lose the most species as they either go extinct or shift ranges. A few species will move in, but most are expected to be closely related to one another and to have similar traits, Voskamp and colleagues reported on 19 July in the *Proceedings of the Royal Society B*.

Northern North America and Eurasia will gain species as birds migrate away from warmer regions. But there, too, many of the newcomers will be closely related to species already present, the data showed. For example, the collared flycatcher (*Ficedula albicollis*) will move farther into northeastern Europe, joining its similar-size cousin, the pied flycatcher (*F. hypoleuca*).

Taken together, the papers forecast a more uniform avian world, a sad prospect for birdwatchers as well as a blow to ecosystems. Some evidence suggests homogenization is happening in amphibians and mammals, too. In the future, "It's the more unique species that will be at a disadvantage," Jarzyna says. ■

INFECTIOUS DISEASES

Antibiotic after sex could help curb infections

Doxycycline reduces risk of sexually transmitted diseases—but will it fuel drug resistance?

By **Jon Cohen**

A new study has shown that taking a single dose of a widely used, cheap antibiotic within 3 days after condomless sex can help prevent chlamydia, syphilis, and gonorrhea, three sexually transmitted infections (STIs) that have soared in the United States, Europe, and elsewhere over the past 2 decades.

The study, mainly in men who have sex with men (MSM) in San Francisco and Seattle, was halted in May after an independent data monitoring board found that the strategy, known as doxycycline post-exposure prophylaxis (doxyPEP), reduced the risk of chlamydia and gonorrhea by more than 60%—a result so convincing there was no need to continue the study. DoxyPEP also appeared to protect against syphilis, but too few cases occurred during the trial to reach statistical significance. The data were scheduled to be presented this week at the 24th International AIDS Conference in Montreal.

"This is very encouraging," says Carlos del Rio, an HIV/AIDS clinician and researcher at Emory University School of Medicine. But there are worries the regimen could trigger antibiotic resistance in the three bacteria that cause these diseases, and scientists are divided about whether the data warrant doxyPEP's introduction now. "It is still a controversial topic," says Jean-Michel Molina of the University of Paris Cité, who led a similar, smaller study in France a few years ago. "I don't think we know enough to recommend the strategy yet."

Doxycycline, a relative of the antibiotic tetracycline, has been around for more than 45 years and is commonly used to treat and prevent acne and Lyme disease. It also acts on parasites and is widely prescribed to prevent malaria in travelers. The drug typically has few side effects beyond stomach upset and increased sensitivity to sunlight.

The idea for the new doxycycline study comes from a small trial published in 2015. Fifteen MSM, who were at high risk of STIs because of their sexual behavior and already took medications for their HIV infections, added the antibiotic to their daily pill regimen. When compared at 48 weeks with 15 similar MSM not given the antibiotic, the men who took doxycycline this way—a form of pre-exposure prophylaxis (PrEP)—had significantly fewer infections with all three bacterial STIs. "That was promising—as a pilot study," says Jeffrey Klausner, an infectious disease specialist at the University of Southern California who co-led the trial.

Molina's study in France, published in 2018, was based on the same idea, but with a slight difference: Doxycycline was used as PEP, not PrEP. Participants took the drug once, within 72 hours *after* having condomless sex. The team found this reduced the risk of acquiring chlamydia or syphilis by 70% and 73%, respectively, but there was no effect on gonorrhea rates.

Clinician Annie Luetkemeyer examines a study participant in San Francisco.

PHOTO: JOSH EDELSON

The new study enrolled 544 participants—mostly MSM, but also some transgender women and gender diverse people—deemed at high risk for STIs. Two-thirds of them were asked to follow a protocol similar to the one in the French study. The remaining participants received standard STI testing and treatment. Everyone knew what group they were in. "We really wanted to do a study in a fairly real-world setting," says one of the principal investigators, Annie Luetkemeyer, an infectious disease clinician at the University of California, San Francisco.

Leandro Mena, who heads the division of the U.S. Centers for Disease Control and Prevention (CDC) that oversees STIs, says he is "very excited" about the findings. "We want to have the opportunity to fully review the data," Lena says, but CDC may soon issue preliminary guidance about the regimen for high-risk populations.

Researchers differ on whether doxyPEP's benefits outweigh the risk of driving antibiotic resistance. *Neisseria gonorrhoeae*, unlike the bacteria that cause the other two STIs, readily develops resistance. DoxyPEP will make that worse, Molina predicts—"That's a no-brainer." What will happen with chlamydia and syphilis, which have not yet evolved to resist antibiotics, "we don't know—we hope," he says.

The U.S. team is still analyzing antibiotic resistance data from its study. But even if doxyPEP increases resistance in *N. gonorrhoeae*, it won't have a practical impact because a different antibiotic class is now used to treat gonorrhea, notes Connie Celum, an epidemiologist at the University of Washington, Seattle, and one of the principal investigators. Besides, she says, doxyPEP would only be recommended for people at high risk of STIs, a relatively small group.

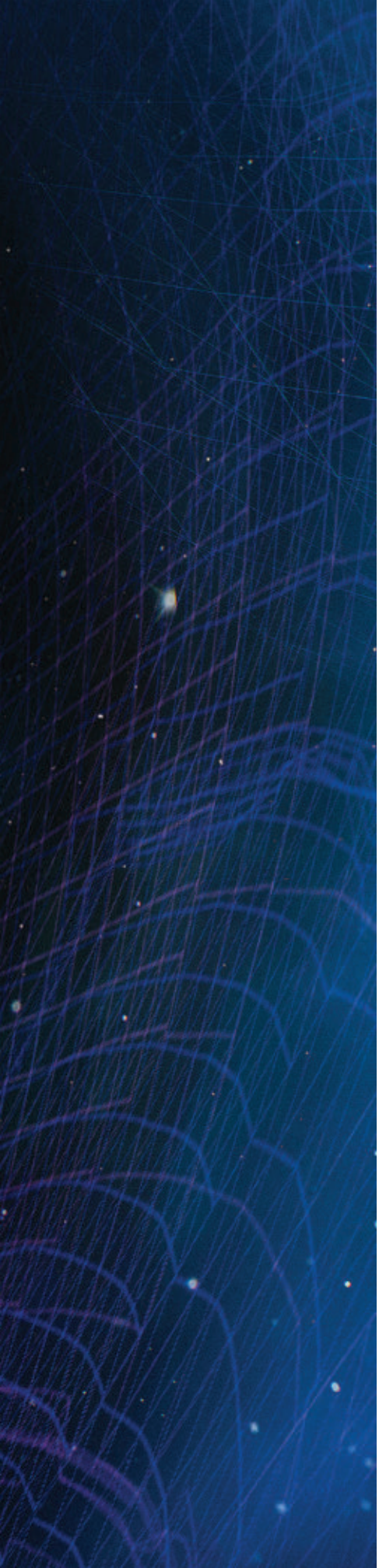
Celum is also involved in a doxyPEP study in Kenyan women who take PrEP for HIV and have disproportionately high rates of STIs. "We need interventions now, and this is a safe, inexpensive, well-tolerated drug that really seems to have few downsides," she says. And treating those at highest risk may help reduce the spread of the STIs in the wider population, Celum adds.

The strategy may have other downsides, however, notes Christopher Fairley, who heads the Melbourne Sexual Health Centre at Monash University, Clayton. Antibiotics can disrupt the bacterial microbiome in the gut, for example, and resistance genes can jump between bacteria. Many gonorrhea and chlamydia infections cause no symptoms, resolve without treatment, and are "of no great significance," he adds. (Untreated syphilis, in contrast, can damage the eyes, ears, and brain, and can lead to miscarriage and stillbirth in pregnant people.)

But Klausner says the benefits outweigh the risks. Some physicians who treat MSM already prescribe doxycycline for STI protection, he notes. "Patients in the community know about it," he says. "The time is now to come up with recommendations." ■

THE LONG SHOT

The Voyager probes raised perplexing questions as they exited the Solar System. Now, scientists have conceived new missions to interstellar space



CREDITS: (ILLUSTRATION) JOHNS HOPKINS APL/MIKE YAKOVLEV; (PHOTO) MARVIN JOSEPH/*THE WASHINGTON POST* VIA GETTY IMAGES

By **Richard Stone**

Before embarking on his Ph.D., Ralph McNutt had never been east of the Mississippi River. But soon after the young Texan arrived at the Massachusetts Institute of Technology (MIT) in the fall of 1975, he found himself on a voyage to the edge of the Solar System—and beyond. Casting around for a research assistantship, he ended up in the office of plasma physicist Herbert Bridge, a towering figure in space science who had overseen the cloak-and-dagger effort to dismantle and ship Harvard University's cyclotron to New Mexico for the Manhattan Project during World War II. Bridge evidently saw a familiar spark in McNutt and invited him to work on a plasma detector for Voyager, the epic mission to the outer planets that began in 1977. "I said, 'Where do I sign up before you change your mind?'"

Can Ralph McNutt rally support for a mission he's mused over for half a century?

Now, this veteran of Voyager, one of NASA's greatest scientific triumphs, wants to wheel his own passion project onto the launchpad. McNutt and colleagues at the Johns Hopkins University Applied Physics Laboratory (APL) have laid out a concept for Interstellar Probe (IP), a $3.1 billion mission to pick up a scientific gauntlet that the two Voyager probes threw down a decade ago after leaving the heliosphere, the Sun's zone of influence. Few expected the spacecraft to survive that long, yet their beguiling observations, still trickling in, have upended many beliefs about the Solar System's outer limits. "A lot of our preconceived notions didn't work out too well," McNutt says.

The Voyager data are so mystifying that some prominent researchers assert the probes haven't made it to interstellar space yet, perhaps because the bounds of the heliosphere stretch farther than generally thought. Gazing out from Earth's perch won't settle the matter. "The only way to see what our fishbowl looks like is to be outside looking in," McNutt says. "We need to get modern instruments out there," adds Lennard Fisk, a space physicist at the University of Michigan (UM), Ann Arbor. "In that sense, Interstellar Probe would be revolutionary."

Now, McNutt needs to convince a jury of his peers. His team has delivered a concept study of IP to the decadal survey of solar and space physics, a community exercise led by the National Academies of Sciences, Engineering, and Medicine that will set the field's priorities for the next 10 years. The panel is set to begin deliberating next month and deliver its verdict in 2024. A thumbs-up for IP would go a long way toward securing NASA support for a probe that would, ideally, lift off in 2036. The timing would allow it to rendezvous with Jupiter and its potent gravity, which would sling the probe toward interstellar space. It would arrive about 16 years later, in half the time it took Voyager.

Chinese scientists are designing a similar mission, called Interstellar Express, that could launch around the same time. Buckle up, enthuses Jim Bell, a planetary scientist at Arizona State University, Tempe, and past president of the Planetary Society. "It's a space race to the edge of the Solar System!"

One challenge for McNutt and his colleagues is selling a mission expected to last at least 50 years, requiring three or more generations of scientists. More daunting may be winning hearts and minds in space physics, which is dominated by experts on space weather—the solar flares and coronal mass ejections that can wreak havoc on satellites and power grids. "People are overly scared that one big project will suck out all the funding for the rest of the science we want to do," says APL space physicist Pontus Brandt, chief scientist on the IP mission concept study. But Merav Opher, an astrophysicist at Boston University, says broadening the field's boundaries is important. "It's myopic if we continue to fund just space weather."

"Voyager on steroids," as McNutt calls IP, may stumble at this first hurdle. "It's a long shot," says Bell, who doesn't have a stake in the project. But IP has a powerful champion in McNutt, says Opher, who calls him "a fantastic mover and shaker." According to Bell, McNutt's superb mentoring abilities will also be critical. "You really do have to think beyond your own lifetime," he says.

INTERSTELLAR SPACE has been a lifelong pursuit for McNutt, who says he was "an introverted and nerdy kid" with a passion for science fiction. One work that left a deep impression was Robert Hein-

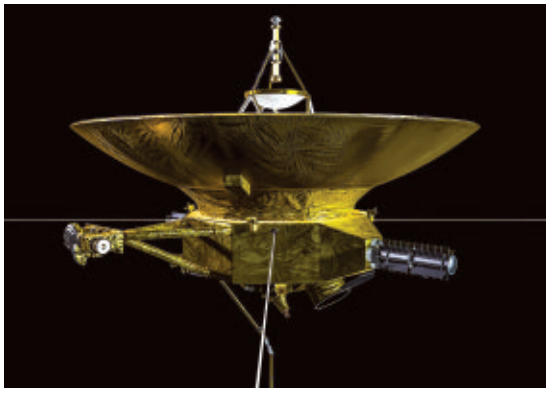
lein's *Time for the Stars*, whose premise was the twin paradox, an early 20th century thought experiment that sought to explain a mind-bending aspect of Albert Einstein's special theory of relativity. In the novel, a telepathic teenager joins an expedition to search for habitable planets around other stars; he spends 4 years on a spaceship that can travel at close to light-speed. He returns home to find his earthbound identical twin had aged 71 years. That premise inspired McNutt, age 16, to concoct an interstellar mission as his project for the 1970 Fort Worth, Texas, science fair. He laid out the physics hurdles of such an epic voyage and even crafted a prototype spacecraft from poster board, balsa wood, and Elmer's glue.

In high school, McNutt struggled to satisfy his thirst for science. School administrators "were more interested in preventing kids from dropping out," but he and several fellow students successfully petitioned for a physics course. A few years later, McNutt got a chance to meet the "father of space travel": Wernher von Braun, a former Nazi rocket scientist who moved to the United States after World War II and became the chief architect of NASA's Moon program. Von Braun was giving a talk at Texas Christian University and McNutt was tapped for a student panel that would pose questions. He asked von Braun whether NASA had plans to put humans on Mars by, say, 1990. That wasn't in the cards, von Braun responded dryly. Instead, he said, the space agency would focus on robotic probes. "I was really irritated," McNutt says. "I was thinking something like, 'What the hell's wrong with you?'"

McNutt came away from the encounter with von Braun's autograph—now in his basement, along with the science fair model—and a fierce determination to become a space scientist. He had a knack for math—"I used to do slide rule speed competitions," he confesses—and majored in physics at Texas A&M University, College Station. At MIT, as a junior member of the Voyager team, he got to go to Cape Canaveral for Voyager 1's launch in 1977, and he vividly remembers a visit 2 years later to mission control at the Jet Propulsion Laboratory (JPL). TV monitors in JPL's cafeteria were showing the first images of Io, Jupiter's flamboyantly colorful volcanic moon. "It looked like a rotting orange or a pizza pie. I thought, 'Oh my God, it's so beautiful.'"

Voyager's revelations about the enigmatic outer planets kept coming. And the dauntless probes kept going. By the early 2000s, it seemed plausible that one or both would reach the heliopause—the boundary between the heliosphere and interstellar space, where the Sun's blast of charged particles, the solar wind, peters out. Tempering that thrilling prospect was the fact that the probes were engineered primarily to interrogate Jupiter's powerful magnetosphere, not the far weaker fields and particles of the interstellar medium. "By today's standards, the information you can get from the Voyager spacecraft is primitive," Bell says. Still, McNutt adds, "The fact that we could get something was a whole lot better than nothing."

A BIG SURPRISE came in 2007, when Voyager 2, diving below the ecliptic plane in which the planets orbit, crossed the termination shock: where the solar wind first starts to falter as it is buffeted by the interstellar gas and dust the Solar System is barreling through. Voyager 1 had crossed the shock 3 years earlier, some 94 astronomical units (AU) from Earth. (One AU, the average distance between Earth and the Sun, is approximately 150 million kilometers.) But its plasma detector had failed at Saturn in 1980, so it could not measure the slowing of the solar wind. Models had predicted the wind would decelerate from 1.2 million kilometers per hour to about 300,000 kilometers per hour. But Voyager 2 clocked a windspeed of 540,000 kilometers per hour. "Going through the termination shock, people said, 'WTF?'" Brandt says.



Interstellar Probe would fly twice as fast as Voyager.

Also mystifying was that Voyager 2 crossed the shock a full 10 AU closer to Earth than Voyager 1. After a Voyager team member broke the news at a conference in Switzerland, "Everybody was like, 'What's going on?' says APL's Elena Provornikova, IP's heliophysics lead, who was then at the Russian Academy of Sciences's Space Research Institute in Moscow. "We immediately started talking about what could cause this asymmetry—what could be the physics behind it."

Space physicists later worked out that models had largely ignored interstellar magnetic fields, which compress the heliosphere below the ecliptic, Provornikova says. The models also assumed the solar wind is a steady gale. But it fluctuates with the Sun's 11-year cycle of magnetic activity, another reason why the two probes reached the shock at different distances.

To explain the weakness of the termination shock, space physicists turned to findings from other planetary probes such as Cassini, the spacecraft that peeled away the mysteries of Saturn and its rings. One was a better understanding of "pickup" ions: neutral atoms, primarily hydrogen, from interstellar space that become ionized when they encounter the solar wind or the Sun's ultraviolet radiation. "Voyager was not equipped to measure pickup ions," Brandt says. "And these are really central here." Scientists inferred that pickup ions riding along with the solar wind would gain enough energy crossing the termination shock to explain why the wind did not slacken as much as predicted.

After navigating that first boundary, the Voyager probes entered the heliosheath, the region where the diminished solar wind continues to wither under a fusillade of gas and dust as the Solar System plows through space. Before Voyager's encounter, the heliosheath was viewed as the heliosphere's thin "skin." But with a stronger solar wind emerging from a weak termination shock, the sheath ought to be thicker. The solar wind would run farther before sputtering to a stop at the heliopause, where the hot, wispy plasma of our heliosphere gives way to the cold, dense plasma of interstellar space.

Without a working plasma detector, Voyager 1 was hard-pressed to confirm that picture. But in early 2013, mission scientists, sifting data from other detectors, declared that the probe had in fact left the heliosphere months earlier, on 25 August 2012—some 122 AU from Earth. A precipitous drop in higher energy solar wind ions and a concomitant rise in cosmic rays clinched the case, they said. Six years later, Voyager 2 hit the heliopause at nearly the same distance from the Sun, in a different phase of the solar cycle—suggesting that unlike the termination shock, the heliopause is insensitive to solar variation. "That was just incredible," Provornikova says.

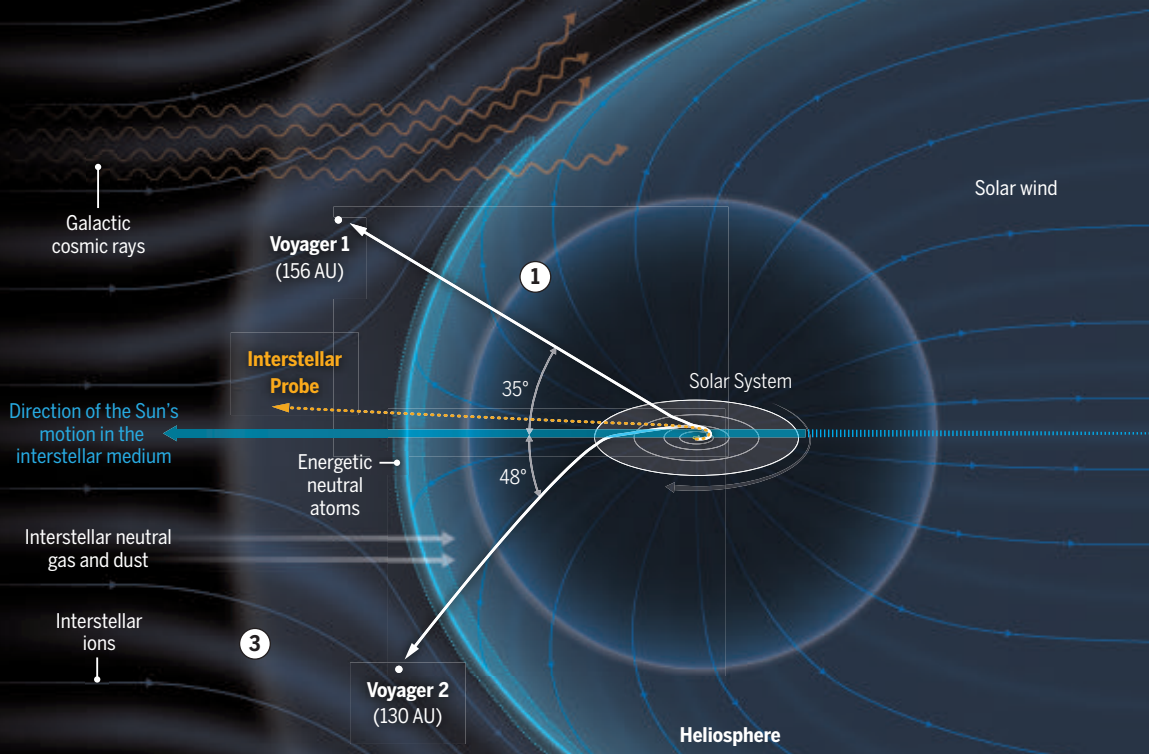
Other data didn't add up. The Sun's magnetic field, embedded in the solar wind, is twisted into a spiral by the Sun's rotation. Crossing the heliopause, Voyager 1 should have observed a change in the direction of the magnetic field, as the solar wind's twisting field gives way to differently oriented interstellar fields. "But it was basically the same damn direction from the Sun," Brandt says. "All the people that know the theory behind it, they're perplexed."

CREDITS: (IMAGE) JOHNS HOPKINS APL/STEVE GRIBBEN; (GRAPHIC, OPPOSITE PAGE) C. BICKEL/*SCIENCE*

Mission of a lifetime

Researchers are pushing NASA to build the $3.1 billion Interstellar Probe (IP), which would explore the plasma environment of the Sun and interstellar space. It would follow in the footsteps of Voyager 1 and 2, which NASA launched in 1977. By most accounts, the twin probes crossed the heliopause into interstellar space a decade ago. If launched by a heavy lift rocket in 2036, IP could leave the heliosphere by 2052, in less than half the time it took Voyager.

Graphic by **Chris Bickel**



A solar cocoon

The heliosphere is the Sun's zone of influence, a bubble inflated by charged particles in the solar wind.

1 Termination shock
Buffeted by neutral gas and dust that penetrate the heliosphere, the solar wind drops from supersonic to subsonic speeds here.

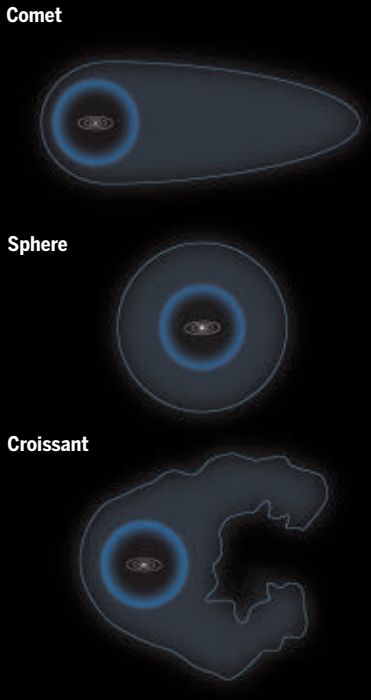
2 Heliopause
At this boundary, the solar wind is stopped by plasma in the interstellar medium. Cosmic rays and energetic neutral atoms should become more prevalent. Magnetic field directions should also shift, although this has not been detected.

3 Bow shock
When the interstellar wind slows suddenly as it nears the heliopause, a shock wave analogous to the termination shock can form. But some studies suggest the Solar System is not moving fast enough to create one.

*One astronomical unit (AU) is 150 million kilometers, the average distance between the Sun and Earth.

Shape shifting

In the classic picture, the heliosphere is shaped like a comet—squashed in the direction of the Sun's orbit around the Galaxy. But data from NASA's Cassini mission point to a spherical shape, whereas other models suggest it may be lobed, like a croissant.

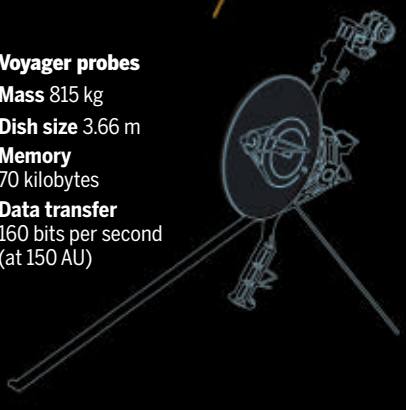


Voyager on steroids

Like Voyager, IP would measure magnetic fields, plasma flux, and cosmic rays. It would also bring new tools such as a dust analyzer. Transmissions would be much faster. "It's like comparing a rotary telephone from 1935 to an iPhone 13," says project leader Ralph McNutt.

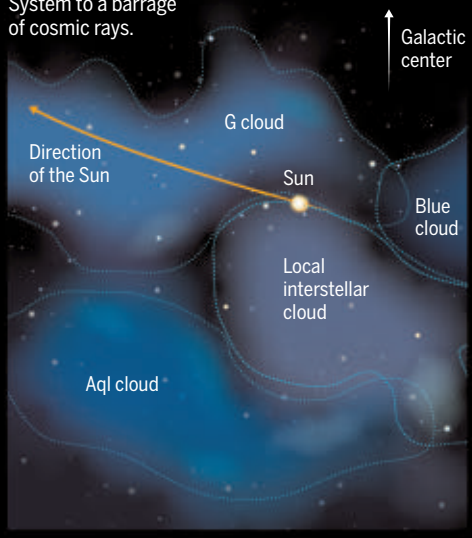
Interstellar Probe
Mass 860 kg
Dish size 5 m
Memory 2 terabytes
Data transfer 16,000 bits per second (at 150 AU)

Voyager probes
Mass 815 kg
Dish size 3.66 m
Memory 70 kilobytes
Data transfer 160 bits per second (at 150 AU)



A cloudy future

Sometime in the next 2000 years, the Sun is expected to leave a local cloud of gas and dust. After that, it may enter the G cloud. If it is denser, it would squash the protective magnetic bubble of the heliosphere, potentially exposing the Solar System to a barrage of cosmic rays.



Fisk thinks it's a sign the probes haven't yet reached interstellar space. In the 1 March issue of *The Astrophysical Journal*, he and UM colleague George Gloeckler propose Voyager 1 and 2 are still in the heliosheath, where they have encountered a unique plasma containing two magnetic fields: not just the field embedded in the wind, but an additional one created by mobile ions that aren't swept up into the wind. "The physics changes dramatically when you account for that," he says. Provornikova and others hold firm that the probes are in interstellar space, arguing that the solar wind's magnetic field evidently dissipates over much greater distances than earlier models augured. "I don't see a scenario where Voyager's still inside the heliosphere," Opher says.

NO MATTER WHO'S RIGHT, scientists find Voyager's interstellar space oddities irresistible. "Voyager didn't give us the answers we're looking for, and we should capitalize on that," Fisk says. McNutt agrees, and in 2017 he assembled a 45-person team of collaborators—including Fisk and Opher—to flesh out a mission concept. Scientists have contemplated an interstellar mission for 50 years, since before Voyager, McNutt says, but "nobody had sat down and run the numbers and done the engineering." The group released its 498-page report at the American Geophysical Union meeting in December 2021.

The mission concept study decisively settles one engineering question: whether to launch the probe toward the Sun and use its enormous gravity as a slingshot, an idea called the Oberth maneuver. After huddling with experts at a thermal materials firm, the IP team worked out that the heat shield needed for the probe to pass so close to the Sun would add too much mass—and risk. "You wouldn't get there any faster" compared with a conventional launch of a lighter probe, McNutt says—provided the launch is on a heavy lift rocket with a rare third and fourth stage. McNutt is eyeing NASA's Space Launch System, a mammoth rocket, bigger than the Saturn V, that could see its first launch this summer as NASA contemplates sending astronauts back to the Moon. And he has put out feelers to SpaceX and Blue Origin about a ride on one of the big launchers those private companies are developing. After the gravity boost from Jupiter, IP should peak at speeds of more than 7 AU per year, about twice as fast as the Voyager probes.

China's Interstellar Express mission would send two probes in opposite directions: one toward the heliosphere's nose, where modelers think it is squashed by the oncoming wind of particles in interstellar space, and the other toward its tail. Observations from both missions "will give us a more comprehensive picture of the heliosphere," says Wang Chi, director-general of the National Space Science Center of the Chinese Academy of Sciences. When his team proposed the mission in 2014, it envisioned a third probe that would launch on a path perpendicular to the ecliptic plane, using nuclear propulsion to escape the heliosphere. But the technical challenges are daunting, and that probe for now is on ice. "As the old Chinese saying goes, a journey of a thousand miles begins with a single step," Wang says. "We should make the two probes successful first."

Voyager 2 being packaged up prior to launch in 1977. It would cross the heliopause 41 years later.

Although McNutt says the Chinese team "holds their cards close to their vest," he, too, views the missions as complementary. "The more the merrier!" he says. "To the extent you'll get different cuts through the heliospheric structure and the nearby interstellar medium, you'll learn a lot about what's going on out there."

APL's concept report lays out a smorgasbord of science IP could tackle, depending on the instruments it carries. High on the list is a suite of four detectors that would measure particles across a broad energy spectrum—from the chilliest plasma and feeblest pickup ions on up to cosmic rays hot enough to fry a DNA strand. "With Voyager we have giant energy gaps," says APL physicist Alice Cocoros. Better detection of pickup ions may be the most important capability, Brandt says, with space physicists just beginning to appreciate the unheralded role they play at the heliosphere's edges.

A dust detector would cure another Voyager blind spot. "We know pretty much nothing about how much interstellar dust actually gets into the Solar System," Provornikova says, or how it interacts with the solar wind. On its outward journey IP could also map the cloud of dust in the outer reaches of the Solar System, leftover from its formation. The contours of this "zodiacal" dust could refine formation models, but they are largely unknown because measurements have only been taken from inside the cloud, McNutt says.

Legging it out beyond the zodiacal cloud would offer another perk: an unobscured view of the extragalactic background light (EBL)—the sum of all radiation produced since the big bang. The New Horizons spacecraft, cruising out beyond Pluto, recently uncovered a mystery when it observed a patch of dark sky and recorded about twice as much visible light as the current census of galaxies can explain, the mission team reported in the 1 March issue of *The Astrophysical Journal*. Equipped with the right instruments, McNutt says, IP "could for the very first time determine the EBL's absolute brightness" across all wavelengths.

Once in interstellar space, IP could also follow in the tradition of other far-ranging probes and look back toward home. But instead of the pale blue dot of Earth, it would capture an image of the entire heliosphere. "You can solve this in one fell swoop," McNutt says, using a unique camera that images a swirling Van Gogh–esque nightscape of energetic neutral atoms (ENAs) generated in the heliosheath when solar wind ions collide with interstellar hydrogen atoms, neutralizing the ions. High-energy ENAs—those above 50 kiloelectron volts (keV)—are especially revealing. "Simulations show that once you get above 50 keV, something remarkable happens—you start seeing images of the shape of the heliosphere," Brandt says. But he says with a grin, "It will probably be the most expensive picture in history."

NASA got a preview in 2008, when it put a spacecraft the size of a bus tire called Interstellar Boundary Explorer (IBEX) in orbit around Earth. Its two ENA cameras captured the first all-sky map of ENAs in the outer heliosphere—and they revealed a stunner: a winding ribbon that is richer in ENAs than surrounding areas. "The great circle in the sky," as Brandt calls it, may be a region just beyond the heliopause where ions snared in a

PHOTO: NASA/INTERIM ARCHIVES/GETTY IMAGES

magnetic field spawn ENAs. IP's ideal launch in 2036 would shoot the probe right through the ribbon.

Early IBEX data supported the traditional notion of a comet-shaped heliosphere, with a tail extending two to three times farther into space than the nose. But subsequent measurements from IBEX, Cassini, and Voyager point to a more rounded heliosphere, and recent modeling suggest it is concave on one side, like a croissant. As an encore to IBEX, NASA in 2025 plans to launch the Interstellar Mapping and Acceleration Probe (IMAP), which would peer out at the heliopause from an orbital station between the Sun and Earth with much finer imaging resolution. "IMAP will bring a lot to the party," McNutt says. But IP, he says, will be able to provide the most revealing ENA map of all once it exits the heliosphere and snaps that picture of high-energy ENAs lighting up the heliosheath.

THE SCIENCE wouldn't stop after the probe reaches interstellar space. Opher says IP would be a "game changer" in our understanding of the interstellar clouds of gas and dust the Solar System hurtles through during the Sun's 230-million-year orbit around the center of the Milky Way. Like oases in a desert, these clouds are likely remnants of stellar nurseries, rich wellsprings of hydrogen that collapsed under gravity to form stars. Space physicists have put together a rudimentary cloud atlas of the local interstellar neighborhood. "It looks like a child's sketch—but that's all we've got," Brandt says.

IP would directly sample gas, dust, and other properties of the Local Interstellar Cloud, the Solar System's home for the past 60,000 years. And by measuring the absorption of starlight by dust and hydrogen atoms, it could probe the nearby G cloud, into which we will plunge in the next 2000 years—if the transition hasn't already begun. "We have no clue what's going to happen next," Brandt says. The denser and colder a cloud is, the more momentum it will sap from the solar wind. That could squash the Sun's magnetic cocoon, to our biosphere's detriment.

Voyager found that 75% of the cosmic rays heading our way from interstellar space get filtered out in the heliosphere's outer reaches. If the encounter with the next cloud squeezes the heliosphere all the way down to Earth's orbit, life forms would be exposed to an intense radiation environment that would riddle DNA with mutations, Brandt says.

There's evidence of such an event around the time early hominids were just beginning to pick up stone tools, and Brandt muses on a possible connection. "Let that creep up your spine for a moment," he says. In recent years, scientists have discovered iron-60 isotopes in ocean crust samples dating from 2 million to 3 million years ago (*Science*, 16 July 2021, p. 269). Iron-60 is not found naturally on Earth: It's forged in the cores of large stars. So, either a nearby supernova blasted the heliosphere with the iron dust, or the heliosphere drifted through a dense cloud laden with iron-60 from a previous supernova. Either way, Brandt says, "The heliosphere was way in, and we had a full blast of galactic cosmic rays and interstellar matter for a long, long time." To look for relics of other such events, IP could use plasma wave antennas to essentially take the temperature of nearby electrons. Hot regions might mark the blast paths of material from past supernovae.

Mission control at NASA's Jet Propulsion Laboratory at the time of Voyager 1's encounter with Saturn in 1980.

The IP team is thinking big in other ways—even the possibility that the probe will ultimately stray near another star and fall into alien hands. Each Voyager probe carries a golden record filled with music and voices sampling Earth's cultures. IP would likely bear an updated digital version: a thumb drive, perhaps, offering a flavor of life on Earth for aliens cosseted in their own heliospheres—as long as an extraterrestrial IT department can figure out how to read it.

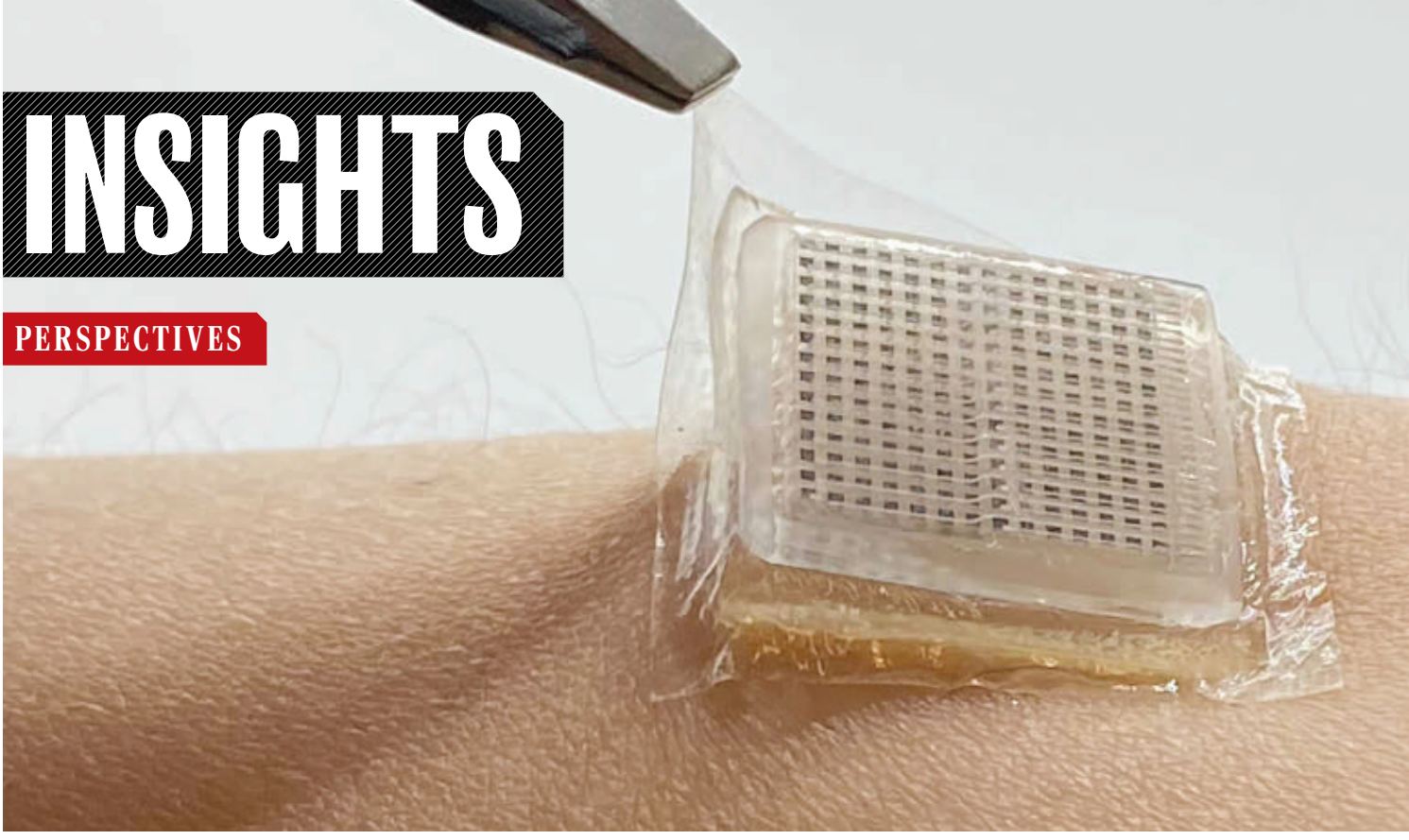
IF THE DECADAL SURVEY endorses IP and NASA embraces it, the space agency would need to persuade Congress that the Star Trekkian emissary is worth the price tag—and then decide which lab would honcho it. Fresh off APL's latest success—the $1.5 billion Parker Solar Probe, which is flying closer to the Sun than any mission ever—McNutt's team is eager to put their sales hats back on. "We have a lot of missions under our belt. Parker came in around $100 million underbudget," Brandt says.

In the meantime, APL is assiduously breeding the next generation of IP scientists. Literally. "We counted 13 babies born during the concept design study," says Cocoros, who has high hopes her son Luke, who's about to turn 2, will become enamored with space. She's expecting a daughter in November. "I guess she makes 14!" During the design study, Cocoros served as a bridge between scientists and engineers as they strove for a sweet spot: an instrument payload that would meet key science objectives without growing too fat for a conventional launcher. "I loved being the glue in the middle of the project," she says.

Cocoros sees McNutt as a mentor. "I adore him. If you ask him a question, he tells you a story. A novel," she says. She says his wealth of knowledge is mirrored in his APL office, which is cluttered with the memorabilia of a life devoted to space. "It's like his brain. Piles and piles of VHS tapes of past missions, giant binders of I don't know what."

With future IP leaders waiting in the wings—and some still in diapers—McNutt hopes his push for the stars will at last get the go-ahead. "We don't want to kick this can down the road farther than it already has been," he says. For this Voyager veteran, after all, it's a trip that's been a half-century in the making. ■

PHOTO: PETER RYAN/SCIENCE SOURCE



MEDICAL IMAGING

Seeing inside a body in motion

Adhesive ultrasound patches can provide medical imaging for patients on the go

By **Philip Tan**[1] *and* **Nanshu Lu**[1,2,3,4,5]

Ultrasonic imaging is one of the most powerful and commonplace medical tools for noninvasive visualization of soft tissues inside the body. However, ultrasonography requires highly trained sonographers to position and orient the transducer on the surface of the patient's body, and image quality is highly dependent on the steadiness of the operator's hand. Because of this, ultrasonic imaging has been largely limited to short and static sessions not only for logistical reasons but also because of the very real threat of musculoskeletal injuries for sonographers from repetitive motions associated with transducer manipulation (*1*). Coincidentally, there has also been a sonographer shortage worldwide over the past decade, and the demanding and specialized training required for certifying sonographers does not help alleviate this problem (*2*). On page 517 of this issue, Wang *et al.* (*3*) introduce a bioadhesive ultrasound (BAUS) patch with the potential to overcome many of these outstanding challenges.

Over the past few years, several wearable ultrasound patches have been developed (*4*–*7*). Although there have been improvements in transducer thinness and softness, none satisfy all the patch-tissue-interface requirements for high-quality ultrasonic imaging. To be ready for real-world applications, the patch must have stable adhesion to the body, the acoustic energy from the transducer must be transmitted efficiently into the body, and the device must not cause skin irritation. In a standard ultrasound exam, liquid gel is the de facto couplant, that is, a material that improves acoustic transmission at the transducer-tissue interface. However, conventional ultrasonic gel is only suitable for short imaging sessions because it dries out or drips down over time. Conversely, elastomer (*4*, *5*, *8*) and hydrogel (*8*) couplants have been explored but tend to have less-than-desirable transmissivity or usable time, respectively.

To overcome these weaknesses, Wang *et al.* combined the strengths of both elastomers and hydrogels by encapsulating a soft hydrogel in a thin elastomer membrane and coating it with a bioadhesive material. This hydrogel-elastomer hybrid, which they named the BAUS couplant, enables more than 48 hours of adhesion to the skin and sustained transmissivity. The soft and adhesive BAUS couplant also alleviates the stiffness of a rigid transducer array, making it more comfortable to wear. The rigid transducer array allows for a high transducer density, with 400 transducers per square centimeter.

The BAUS patch opens exciting opportunities for increasing accessibility to ultrasound exams. Although, at present, the patch still needs to be attached to a nonportable data acquisition system for processing the images, it could provide immediate relief for sonographers, for example, during sessions when a patient is undergoing anes-

[1]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA. [2]Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, TX 78712, USA. [3]Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA. [4]Texas Materials Institute, The University of Texas at Austin, Austin, TX 78712, USA. [5]Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712, USA. Email: nanshulu@utexas.edu

PHOTO: WANG *ET AL.* (*3*)

Wang *et al*. created a hydrogel-elastomer adhesive that can be used as a conductive medium for an ultrasound transducer, which provides the flexibility needed for a wearable ultrasound imaging device.

thesia and constant monitoring of cerebral blood flow is required (*9*). In this scenario, the BAUS patch can simply be stuck to a patient's temple and provide stable blood flow imaging. The patch also opens opportunities for long-term, dynamic, high-resolution imaging of internal organs. For example, discreet monitoring of bladder volume could help people who suffer from urinary dysfunction. The inability to sense bladder fullness is a cause of urinary incontinence, leading to physiological and psychological injury, and a portable ultrasonic device for measuring bladder fullness can provide private alerts for these individuals to relieve themselves (*10*). For managing cardiovascular disease, stress echocardiograms provide valuable insights into cardiac functions, including valvular regurgitation, valve gradients, chamber volume, and myocardial strain tracking (*11*). However, the difficulty of maintaining the transducer probe during motion puts substantial logistical limits on these exams.

Thanks to excellent coupling to the tissue, the BAUS patch can capture high-quality images during a wide range of motions (see the figure). Muscular imaging enables quantification of muscle contraction, which enhances athletic training and rehabilitation of neuromuscular disorders (*12*). Also, ultrasonic imaging of the lungs has been a powerful tool for screening for respiratory diseases such as COVID-19 (*13*). A "set-and-forget" ultrasonic patch would allow for continuous monitoring of contagious patients with minimal exposure risk to the medical staff.

However, there do exist several limitations of the BAUS patch that will require future improvement. Wang *et al.* introduced a plug-and-play snap connector system that could link to several dozen transducers. However, three-dimensional ultrasonic imaging will require thousands of transducers to provide sufficient resolution. Controlling thousands of transducers requires extensive circuitry and hardware, which will limit maneuverability and mobility. Forming interconnects between transducers and the circuitry becomes more challenging with more transducers involved. Research into "ultrasound on a chip" will offer the much-desired integration of transducers with requisite analog and digital processing circuitry. The BAUS couplant has the potential to improve the long-term wearability of these systems, despite differing transducer array fabrication approaches.

To make ultrasound completely user-friendly for the average patient, artificial intelligence technology may also be needed to assist with placement and beam steering. Researchers have demonstrated an interactive application using a deep-learning model that guides the operator to manipulate the transducer to produce the best image of a target anatomy (*14*). Moving forward, automated image analysis and interpretation will be just as important as image acquisition. The road to providing universal accessibility for mobile ultrasonic imaging will require concerted efforts in diverse fields of engineering, science, and medicine. ■

REFERENCES AND NOTES

1. J. Gremark Simonsen, A. Axmon, C. Nordander, I. Arvidsson, *BMC Musculoskelet. Disord.* **21**, 156 (2020).
2. L. Waring, P. K. Miller, C. Sloane, G. Bolton, *Ultrasound* **26**, 206 (2018).
3. C. Wang *et al.*, *Science* **377**, 517 (2022).
4. J. S. Kenny *et al.*, *Sci. Rep.* **11**, 7780 (2021).
5. V. Pashaei *et al.*, *IEEE Trans. Biomed. Circuits Syst.* **14**, 305 (2020).
6. H. Chen *et al.*, bioRxiv 10.1101/2021.11.09.467971 (2021).
7. C. Wang *et al.*, *Nat. Biomed. Eng.* **5**, 749 (2021).
8. R. Manwar, L. Saint-Martin, K. Avanaki, *Chemosensors* **10**, 181 (2022).
9. I. K. Moppett, R. P. Mahajan, *Br. J. Anaesth.* **93**, 710 (2004).
10. H. Niu *et al.*, in *2011 5th International Conference on Bioinformatics and Biomedical Engineering* (IEEE, 2011).
11. T. H. Marwick, in *Echocardiography*, P. Nihoyannopoulos, J. Kisslo, Eds. (Springer, 2018), pp. 325–350.
12. B. Van Hooren, P. Teratsias, E. F. Hodson-Tole, *J. Appl. Physiol.* **128**, 978 (2020).
13. C. Xing *et al.*, *Crit. Care* **24**, 174 (2020).
14. J. M. Rothberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019339118 (2021).

ACKNOWLEDGMENTS

P.T. acknowledges the US National Science Foundation (NSF) Graduate Research Fellowship Program. N.L. is supported by the US NSF Addressing Systems Challenges through Engineering Teams (ASCENT) program under grant 2133106.

10.1126/science.adc8732

CANCER

Targeting brain cancer

A new drug is designed to hit an old target with increased precision

By **Roger Reddel** *and* **Adel Aref**

Glioblastoma (GBM) is the most common form of malignant tumor originating in the brain and is one of the most lethal cancers, with only 1 in 20 patients surviving for 5 years after diagnosis (*1*). The relentless growth of cancers such as GBMs results from DNA mutations and epigenetic changes, but these also create opportunities for targeted therapies. On page 502 of this issue, Lin *et al.* (*2*) describe a new approach to killing GBM cells that targets a well-known molecular change in GBMs: the decreased expression of the DNA repair enzyme O^6-methylguanine methyltransferase (MGMT). The authors develop a new agent that exploits this reduced DNA repair capacity and propose that their approach could be generalized to other cancers with defective DNA repair.

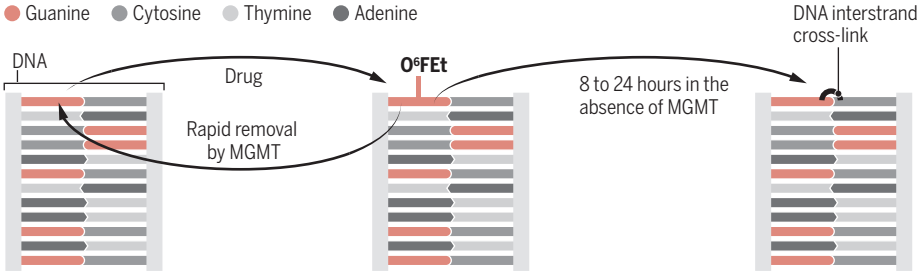
The expression of the *MGMT* gene is repressed by DNA methylation of its promoter. This occurs in ~50% of GBMs and >70% of lower-grade gliomas (grades II and III), and less frequently in other cancers (*3*). How this change contributes to oncogenesis is unclear, but MGMT interacts with the p21 and proliferating cell nuclear antigen (PCNA) cell cycle proteins (*4*), so lack of MGMT protein presumably contributes to unrestrained cell proliferation. What is much clearer is that MGMT provides a critical function in maintaining the integrity of DNA, with its primary function being to repair O^6-methylguanine DNA adducts, which cause base-pairing with thymine rather than cytosine and thus can introduce mutations during DNA replication. MGMT repairs these adducts through a "suicide" mechanism in which it transfers the methyl group from damaged DNA to the sulfur group in its internal acceptor site, inactivating itself (*3*).

One of the few therapeutic interventions that has well-documented benefit in GBM

Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, NSW, Australia. Email: rreddel@cmri.org.au

Inducing toxic DNA lesions in glioblastoma

The drug KL-50 was designed to deposit an O^6-2-fluoroethyl (O^6FEt) lesion on guanine bases that can be rapidly removed by the O^6-methylguanine methyltransferase (MGMT) enzyme. If not removed because of MGMT deficiency in a cancer cell, the lesion is slowly converted to a highly deleterious interstrand cross-link, leading to cancer cell death by a mechanism independent of the mismatch repair pathway.



is the combination of radiotherapy and an alkylating drug, temozolomide (TMZ) (*5*). TMZ is a prodrug that converts to an alkylating agent under physiological conditions and causes various types of lesions in DNA, most of which are readily repaired by the base excision repair pathway. A small percentage of the lesions, however, are O^6-methylguanine adducts, so GBMs with low MGMT expression accumulate these TMZ-induced DNA adducts. This leads to cell death through a mechanism that involves the mismatch repair (MMR) pathway (*6*) and is an example of a synthetic lethal interaction between a drug and a genetic or epigenetic change (*7*). MGMT deficiency is therefore a biomarker for response to TMZ in GBM and grade II and III gliomas. By contrast, in other types of tumors, MGMT deficiency is either not predictive of TMZ response, or its predictive capability has not yet been established (*8*). In most MGMT-deficient GBM patients, the initial response to TMZ is eventually followed by tumor recurrence coincident with mutations in genes that encode MMR proteins and the development of a hypermutator phenotype.

Lin *et al*. describe a new approach to exploiting MGMT deficiency in GBM cells (see the figure). They designed agents that create a primary DNA lesion that can be reversed by MGMT, but if this does not occur, the lesion evolves into a toxic DNA interstrand cross-link (ICL). ICL formation occurs at a rate that is slow enough for most or all of the primary lesions to be repaired in normal cells for which the drugs should therefore have low toxicity. ICLs cause cell death in a manner that cannot be evaded by MMR mutations, so this design should also decrease the acquisition of resistance. One of these agents, KL-50, was shown to cause maximal ICLs 8 to 24 hours after treatment and to retain selectivity for MGMT-deficient cancer cells as well as killing MMR-deficient cells for which TMZ was ineffective, confirming that the MMR pathway is not involved. Moreover, KL-50 appeared to have an acceptable toxicity profile in mice xenografted with human GBM cells. There will be many steps required to take this potential therapeutic agent from preclinical studies to human trials. It will be of particular interest to determine whether changes in DNA methylation of the *MGMT* promoter (*9*) emerge as a mechanism of resistance, and whether any of the mechanisms other than MMR mutations that result in TMZ resistance (*10*) can also result in resistance to KL-50.

Mutations or epigenetic changes that result in impairment of one or more DNA repair pathways occur in many different types of cancer, especially as a result of treatment (*11*), and mutation of various DNA repair genes is the most common known cause of hereditary cancer. Lin *et al*. propose that their approach could be used to exploit DNA repair deficits in cancer cells other than MGMT deficiency for therapeutic purposes. The principle would be the same: Modify the agent so that it produces an initial lesion that is rapidly repaired in DNA repair–proficient normal cells—but not when the target DNA repair pathway is deficient—and a secondary toxic lesion that develops slowly.

"...their approach could be used to exploit DNA repair deficits in cancer cells other than O^6-methylguanine methyltransferase deficiency..."

While Lin *et al*. are improving the drugs used to hit a well-established target—MGMT deficiency—the search continues for new treatment targets in GBM. Despite rapidly increasing knowledge about the biology of GBM, profiling of the GBM genomic landscape, and identification of actionable targets (*12*, *13*), clinical outcomes have not improved substantially in the past decade. A network meta-analysis of 40 randomized controlled clinical trials, which included patients with recurrent GBM treated with various regimens (including single agents and combinations of tyrosine kinase inhibitors, chemotherapy, immunotherapy, and antiangiogenic agents), showed no significant difference in overall survival between the different regimens (with overall survival ranging from 3 to 18 months) (*14*).

The reasons that GBM is so difficult to treat are thought to include the extent to which tumor cells infiltrate the surrounding tissue, making surgical cure largely impossible; an intrinsically high level of resistance to radiotherapy and chemotherapy; an immunosuppressive tumor environment; and the extensive heterogeneity among GBMs in different patients and within individual tumors (*15*). Improvements in outcome are likely to come from increased understanding of the tumor biology. The history of the development of successful cancer treatments also suggests that outcomes will be improved by the use of treatment combinations. In this context, combining a potentially improved approach to targeting MGMT deficiency with other therapies such as radiotherapy, another targeted therapy, virolytic agent, and/or immunotherapy could be a promising step forward. ■

REFERENCES AND NOTES

1. S. Lapointe, A. Perry, N. A. Butowski, *Lancet* **392**, 432 (2018).
2. K. Lin *et al*., *Science* **377**, 502 (2022).
3. S. Sharma *et al*., *Anticancer Res.* **29**, 3759 (2009).
4. A. Mostofa, S. R. Punganuru, H. R. Madala, K. S. Srivenugopal, *Neoplasia* **20**, 305 (2018).
5. R. Stupp *et al*., *N. Engl. J. Med.* **352**, 987 (2005).
6. G. M. Li, *Cell Res.* **18**, 85 (2008).
7. A. Ashworth, C. J. Lord, *Nat. Rev. Clin. Oncol.* **15**, 564 (2018).
8. B. I. Hiddinga, P. Pauwels, A. Janssens, J. P. van Meerbeeck, *Lung Cancer* **107**, 91 (2017).
9. J. Feldheim *et al*., *Cancers* **11**, 1837 (2019).
10. N. Singh, A. Miner, L. Hennis, S. Mittal, *Cancer Drug Resist.* **4**, 17 (2021).
11. S. Negrini, V. G. Gorgoulis, T. D. Halazonetis, *Nat. Rev. Mol. Cell Biol.* **11**, 220 (2010).
12. D. S. Nørøxe, H. S. Poulsen, U. Lassen, *ESMO Open* **1**, e000144 (2017).
13. T. F. Cloughesy, W. K. Cavenee, P. S. Mischel, *Annu. Rev. Pathol.* **9**, 1 (2014).
14. A. Schritz *et al*., *Neurooncol. Adv.* **3**, vdab052 (2021).
15. A. P. Patel *et al*., *Science* **344**, 1396 (2014).

ACKNOWLEDGMENTS

The authors are funded by agencies including the Australian Cancer Research Foundation, the New South Wales Government's Ministry of Health, and the Australian Government's Medical Research Future Fund.

10.1126/science.add4839

GRAPHIC: N. DESAI/*SCIENCE*

MOLECULAR BIOLOGY

Molecular recorders to track cellular events

DNA tapes could be used to record dynamic molecular and cellular events in animals

By **Nanami Masuyama[1,2], Naoki Konno[3], Nozomu Yachie[1,4]**

Starting from a fertilized egg, cells proliferate, pass their genomic information to their progeny, and dynamically change their functions to form diverse tissue structures. Throughout development, intracellular and environmental cues trigger patterns of gene expression that govern cell state transitions and produce further cellular and environmental cues, leading cells to self-organize into functional clusters in spatially distinct areas. How can these processes be investigated? High-resolution molecular snapshots of cells can be obtained by various omics technologies, but these require sample destruction, which precludes analyses over time. Live-cell imaging with fluorescent probes can analyze time-course dynamics but is limited to a small number of molecules or cells. These challenges could be overcome by DNA event recording, where molecular and cellular information is progressively stored in synthetic "DNA tapes" in cells within a multicellular organism, and the stored information is accessed by high-throughput DNA sequencing. However, there are technical limitations that need to be overcome.

One of the DNA event recording modalities, dynamic cell lineage recording using CRISPR-Cas9 genome editing, has accelerated developmental cell lineage tracing. In this system, chromosome-embedded synthetic DNA barcode arrays are continuously and randomly mutated by Cas9, which is recruited by the expression of many barcode-targeting guide RNAs (gRNAs); these mutated arrays are progressively inherited by daughter cells (*1*–*3*) (see the figure). Similar to an evolutionary tree reconstruction from genomic sequences of diverse species, the cell lineage of an animal initiated from a fertilized egg can be reconstructed according to the mutation patterns of the barcode sequences in descendant cells at the time of observation using high-throughput DNA sequencing. Such lineage tracing systems have also been linked to single-cell RNA sequencing (scRNA-seq), where individual cell types identified by scRNA-seq can be linked to their historical cell lineage information (*3*).

Today, new technologies are rapidly revising perspectives on developmental biology, such as the origins of the hypoblast and embryo-restricted epiblast (*4*), blood cell differentiation, and the deciphering of cell proliferation dynamics of metastatic cancers (*5*, *6*). Furthermore, ideas to develop other cell-embedded circuits that can convert environmental and cellular signals to genome-editing activities have emerged (*2*). Combined with a cell lineage recorder, such systems would enable the computational reconstruction and mapping of historical cellular events and cell state changes onto a cell lineage. These ideas could lead to the characterization of high-resolution cell lineages, cell state changes, and cell communications during whole mammalian body development, homeostasis, and malfunction. This would deepen knowledge about developmental and cancer biology that could be translated into therapeutic strategies. However, DNA event recording has yet to trigger the next gold rush in biology because there are some missing components. The current molecular event recording methods need to overcome four substantial limitations in information recording capacity, available molecular sensors to capture diverse biological events, the number of cells that can be analyzed, and computational capacity to reconstruct high-content biological history.

Information recording capacity could be increased by extending the total length of the DNA barcodes. For example, several lineage tracing methods have used multiple short DNA barcode units per cell for the random accumulation of CRISPR-induced mutations. To read out the combinatorial mutation patterns in these barcodes in distal genomic loci, constitutively active promoters are introduced to allow transcription, and the mutation information is captured by scRNA-seq. The current methods use either a few copies of long gRNA-targeting barcode arrays (up to 10 barcodes per array, whose size is restricted by short-read DNA sequencing methods) or many copies of short barcode arrays with single gRNA-targeting units. There seems to be no barrier to combining these two approaches to achieve a high-content event recording system (i.e., encoding many long barcode arrays per cell).

Altering the type of information-writing enzyme would also increase the information recording capacity. Most of the current cell lineage tracing systems use wild-type Cas9, which induces double-stranded DNA breaks that are toxic to cells and can also confer large deletions, especially when multiple breaks are introduced simultaneously. Additionally, these large deletions can overwrite acquired mutations and progressively collapse the barcodes. These limitations could be alleviated by using a CRISPR base editor (BE), fusing a catalytically inactive Cas9 to a nucleoside deaminase that introduces single-nucleotide substitutions at the gRNA target sequence with minimal double-stranded DNA breaks. Most BEs confer unidirectional substitutions (e.g., C→T or A→G) (*7*). This could be a downside, given that the lineage recording ability crumbles when all substrate C or A bases in the DNA barcodes are converted and thus depleted. However, the recent development of dual-function BEs has enabled the simultaneous introduction of C→T and A→G substitutions. Using such a dual-function BE that can confer reversible C:G⟷T:A mutations will enable more-accurate lineage tracing than wild-type Cas9 or single-function BEs. In evolutionary biology, phylogenetic trees have been efficiently reconstructed from revertant mutations. BEs have also been demonstrated to record cellular signals but only for a limited number of events in mammalian cells, such as Wnt signaling (*8*).

The CRISPR prime editor (PE), fusing a catalytically inactive Cas9 to reverse transcriptase, has recently been proposed for scalable, high-content event recording. In these genome-editing systems, a PE gRNA (pegRNA) encodes a synthetic short payload sequence and recruits the PE to the target sequence, in which—coupled with endogenous DNA repair machineries—the reverse transcription activity of PE enables the integration of the payload sequence to the target site (*7*). This enables more diverse data encoding per barcode than wild-type Cas9 and BEs. Preliminary data show that the gene expression activities of 300 enhancers can be recorded to a single target DNA locus of a cell population using a library of pegRNAs, each of which is encoded downstream of a specific enhancer element (*9*). PE has also been programmed for a cascade reaction, where the order of complex events can be recorded by arraying event-specific payload sequences

[1]School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada. [2]Institute for Advanced Biosciences, Keio University, Yamagata, Japan. [3]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [4]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. Email: nozomu.yachie@ubc.ca

at a target DNA locus like a ticker tape (*10*). However, these approaches have not yet been implemented in animal models.

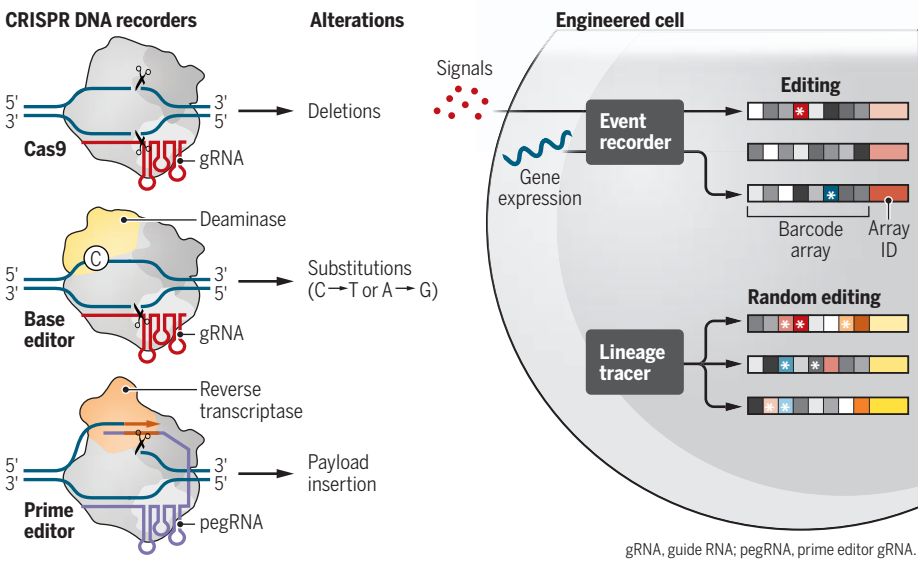
High-content event reconstruction necessitates high-capacity scRNA-seq. A cell lineage recording system that confers at least one unique mutation per cell division would yield the perfect reconstruction of cell lineage if all cells of a body can be analyzed. Analysis of a limited number of cells enables the historical reconstruction of only an upstream cell lineage whose descendant information can be covered by the analytes. The current state-of-the-art scRNA-seq can process multiple millions of cells, whereas a mouse body consists of multiple billions of nucleated cells. Given the rapid development of scRNA-seq since the advent of the droplet-based methods that enabled the analysis of multiple thousands of cells (*11*), it is not unreasonable to predict that technologies that can capture a greater magnitude of cells will appear in the next 5 years. Additionally, spatial RNA-seq has enabled spatial profiling of gene expression patterns in a tissue section or a three-dimensional block in high resolution (*12*). Coupled with DNA event recording systems whose readout information is converted into RNA, this modality also enables the reconstruction of cell lineage and cell state trajectories throughout a physical space. However, spatial RNA-seq allows only the reconstruction of later events that contributed to its formation. Therefore, more-scalable methods need to be developed to capture the whole developmental process of an animal body.

High-performance computing methods also need to be developed with these engineering efforts. Although many of the current cell lineages have been generated by repurposing existing evolutionary tree reconstruction tools, their algorithms need to be optimized for information recording biases in specific lineage tracing methods (*13*). Efficient algorithms for other biological events have yet to be established. Furthermore, no current phylogeny estimation software tool can reconstruct a lineage of more than one million inputs, raising concerns about how to reconstruct a biological history forming a mammalian body of at least hundreds of millions of cells. A new computing framework called deep distributed computing has been proposed to overcome this limitation (*14*). This framework has enabled accurate cell lineage tracing of more than two hundred million mutated sequences. In brief, starting from a single computing node, it reconstructs an upstream lineage and assigns all the input sequences into its subclades, where an independent computing node recursively resolves the next upstream lineage of each subclade. Like a spreading virus, a single computing task autonomously generates many distributed computing tasks, each of which iterates the same procedure. There is no conceptual barrier to implementing other event reconstruction processes into this framework.

Technologies for high-resolution DNA event recording have been maturing, which could expand the horizons of biology in different animal models—perhaps starting from mice for mammalian species. It is expected that this approach will identify unknown cell differentiation behaviors and their underlying molecular and cellular regulatory mechanisms. Analysis of many individuals' developmental processes in a single species in high resolution would characterize deterministic and stochastic cell differentiation patterns and quantify the ability of cells to undergo state transitions, enabling more precise computational modeling of dynamic multicellular systems with biophysical perspectives. Direct observation of human developmental procedures with the same approach is impossible because of ethical considerations regarding the genetic manipulation of human embryos. However, high-resolution developmental maps of diverse nonhuman species and comparative genomics approaches could also reveal human developmental processes. Furthermore, as seen in the emerging CRISPR cell lineage tracing studies of human organoid development (*15*) and tumor progression to metastasis in a mouse model (*5*, *6*), embedding a high-content DNA event recorder system in human pluripotent stem cells for organoid studies and in cancer cells for animal xenograft studies can more directly contribute to understanding their behaviors. Combined with genome-editing technologies that can reconstitute disease mutations in the human genome, these new model systems will also serve as testbeds with substantially expanded data monitoring abilities in therapeutic development for genetic disorders and cancers. ■

The DNA event recording concept

Cell-embedded synthetic CRISPR-Cas information-writing enzymes can progressively record cellular events and lineage transitions into DNA tapes (or barcode arrays) by introducing various types of DNA alteration. DNA tapes are read out by single-cell sequencing, and the biological event histories are computationally reconstructed. Cell lineage can be recorded by continuous accumulation of random mutations into DNA tapes. Other event recording systems necessitate editing event-specific DNA tape positions or inserting specific DNA payloads into the DNA tape.



gRNA, guide RNA; pegRNA, prime editor gRNA.

REFERENCES AND NOTES

1. A. McKenna *et al.*, *Science* **353**, aaf7907 (2016).
2. K. L. Frieda *et al.*, *Nature* **541**, 107 (2017).
3. S. VanHorn, S. A. Morris, *Dev. Cell* **56**, 7 (2021).
4. M. M. Chan *et al.*, *Nature* **570**, 77 (2019).
5. K. P. Simeonov *et al.*, *Cancer Cell* **39**, 1150 (2021).
6. D. Yang *et al.*, *Cell* **185**, 1905 (2022).
7. A. V. Anzalone *et al.*, *Nat. Biotechnol.* **38**, 824 (2020).
8. F. Farzadfard *et al.*, *Mol. Cell* **75**, 769 (2019).
9. W. Chen *et al.*, bioRxiv 10.1101/2021.11.05.467434 (2021).
10. J. Choi *et al.*, *Nature* 10.1038/s41586-022-04922-8 (2022).
11. R. Stark *et al.*, *Nat. Rev. Genet.* **20**, 631 (2019).
12. S. K. Longo *et al.*, *Nat. Rev. Genet.* **22**, 627 (2021).
13. W. Gong *et al.*, *Cell Syst.* **12**, 810 (2021).
14. N. Konno *et al.*, *Nat. Biotechnol.* **40**, 566 (2022).
15. Z. He *et al.*, *Nat. Methods* **19**, 90 (2022).

ACKNOWLEDGMENTS

We thank members of the Yachie laboratory for useful insights, especially S. King and A. Adel. The authors are supported by the Canada Research Chair program, a project grant (PJT 178040) by the Canadian Institutes for Health Research, and a Pilot Innovation Fund (PIF003) by Genome British Columbia. N.M. and N.K. are supported by fellowships from the Japan Society for the Promotion of Science. N.M. and N.K. contributed equally to this work.

10.1126/science.abo3471

GRAPHIC: K. HOLOSKI/*SCIENCE*

MARINE BIOLOGY

Did pollination exist before plants?

Research shows that seaweeds depend on crustaceans for fertilization

PHOTO: © SEBASTIEN COLIN/MAX PLANCK INSTITUTE FOR BIOLOGY/STATION BIOLOGIQUE DE ROSCOFF

Lavaut *et al*. demonstrate the involvement of the crustacean *Idotea balthica* (purple) in the transport of male gametes of the red seaweed *Gracilaria gracilis* (green). This echoes the assisted pollination of flowering plants, which raises questions about the ecological importance and the evolution of these interactions.

By **Jeff Ollerton**[1,2] *and* **Zong-Xin Ren**[1]

The sexual reproduction of seed plants involves the transfer of male gametes—in pollen—to their female gametes. In flowering plants (angiosperms), this is performed with the stigma of flowers, whereas the gymnosperms (such as conifers and cycads) produce a diversity of structures on their reproductive axes to accomplish the same task. This transfer of male gametes is generally known as "pollination" and can be mediated by animals, wind, or water. Animal pollination, principally by insects, is the dominant strategy for angiosperms but also occurs in the extant gymnosperms, as well as some species of mosses. Outside of these plants, no other group of organisms has been demonstrated to interact with animals in this way, until now. On page 528 of this issue, Lavaut *et al.* (*1*) demonstrate that a living species of red seaweed, *Gracilaria gracilis,* uses the isopod crustacean *Idotea balthica* to transfer its male gametes, substantially extending the phylogenetic scope of species that use animals as pollinators.

Although they photosynthesize, seaweeds are a type of algae and therefore are only very distantly related to what is considered true plants. The group to which the red seaweed *G. gracilis* belongs is thought to have evolved around 1 billion years ago, long before plants appeared (*2*). The isopods are not as old, but nonetheless there have been at least 300 million years for relationships to have developed with the seaweed. It is possible that before this, red seaweeds relied on some other now-extinct marine invertebrates as pollinators, but the earliest unambiguous arthropods only appear about 600 million years ago. How these seaweeds were reproducing before this is a mystery, but Lavaut *et al*. suggest that reproduction in the seaweed is "ambophilous"—that is, its male gametes can be transported by the isopods or by water currents. This finding is worth noting because extant seed plants that rely on both wind and animal pollination are widespread, and the strategy is assumed to be underreported (*3*, *4*).

What distinguishes the 6750 described species of Florideophyceae—of which *G. gracilis* is a member—from most other seaweeds is that their male gametes lack a flagellum to provide movement. It was previously thought that they could reach the female reproductive structures passively with water currents, a common strategy used by some aquatic flowering plants (*5*). However, Lavaut *et al*. demonstrate that isopod crustaceans may be attracted to these seaweeds because they provide a habitat in which to shelter from predators and feed on small algae that colonize the surface of the seaweed. In addition to pollinating, grazing by the isopods removes these algae and may increase the growth rate of the seaweed. In this way, the isopods are engaged in a "double mutualism" (*6*) with the seaweed.

Interactions between different species are the engine that makes ecosystems function effectively. Herbivores grazing on plants, predators consuming animals, decomposers breaking down organic matter and recycling it all contribute to the ultimate circular economy of the planet. Ecologists consider the outcomes of these interactions in terms of which species benefit, which are negatively affected, and which have a neutral outcome with respect to their evolutionary fitness—that is, the ability to reproduce. Mutualistic interactions in which both species benefit are often responsible for maintaining populations because of their positive effects on individual reproductive success. Mutualism plays a crucial role in ecology, from gut microbes that facilitate the digestion of cellulose in herbivores, to fish and sea anemones that protect each other from their respective predators, to plant-pollinator relationships (*7*).

Relying on a third party to facilitate reproduction is obviously a vital way for organisms that cannot move around to find mates. Although it was once considered the hallmark of flowering plants, researchers are discovering that this kind of interaction, in which animals bring together different gametes or mating types of sedentary organisms, may be found in taxa that evolved long before angiosperms and gymnosperms, such as mosses (*8*) and fungi (*9*, *10*). The study by Lavaut *et al*. has broadened both the variety and the history of animal-mediated male gamete transfer, taking the concept of pollination from plants to algae

[1]CAS Key Laboratory of Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. [2]Faculty of Arts, Science and Technology, University of Northampton, Northampton, UK. Email: jeff.ollerton@gmail.com; renzongxin@mail.kib.ac.cn

and potentially pushing it back to the earliest evolution of marine invertebrates.

The reliance on an animal to ensure reproduction is not without its risks, especially in a world of rapid anthropogenic change. As Lavaut *et al.* point out, most reproduction in this seaweed occurs in the relatively still waters found in coastal rock pools at low tide. These habitats are threatened by factors such as water pollution, invasive species, physical modification of the coastline, and changing climate and sea level (*11*). In addition, crustaceans in general are susceptible to the increasing ocean acidification related to rising atmospheric CO_2 concentration. Ocean acidity can weaken their exoskeletons through decalcification, although this effect has yet to be thoroughly examined for isopods (*12*). Seaweeds, including *Gracilaria gracilis*, are also ecologically important for the ecosystems in which they occur and are increasingly being touted as a sustainable food source (*13*, *14*). Marine environments are under huge anthropogenic pressure, however, and Lavaut *et al.* demonstrate the importance of preserving not just species within ecosystems but also the interaction networks among different species.

There is so much work still to be done in this area of ecology and discoveries to be made. Despite being the most extensively studied group of plants, recent estimates suggest that only 10% of the more than 300,000 known species of animal-pollinated flowering plants have had their pollinators documented (*15*). There are no doubt many more revelations awaiting the careful observer of species interactions. ■

REFERENCES AND NOTES

1. E. Lavaut *et al.*, *Science* **377**, 528 (2022).
2. E. C. Yang *et al.*, *Sci. Rep.* **6**, 21361 (2016).
3. Y. B. Gong *et al.*, *Plant Species Biol.* **31**, 272 (2016).
4. M. E. Saunders, *Insect Conserv. Divers.* **11**, 13 (2018).
5. P. A. Cox, *Annu. Rev. Ecol. Syst.* **19**, 261 (1988).
6. F. Fuster, C. Kaiser-Bunbury, J. M. Olesen, A. Traveset, *Ecography* **42**, 826 (2019).
7. J. Ollerton, in *Plant-Pollinator Interactions: from Specialization to Generalization*, N. M. Waser, J. Ollerton, Eds. (Univ. Chicago Press, 2006), pp. 411–435.
8. N. Cronberg, R. Natcheva, K. Hedlund, *Science* **313**, 1255 (2006).
9. T. L. Bultman, J. F. White Jr., T. I. Bowdish, A. M. Welch, *Mycol. Res.* **102**, 235 (1998).
10. B. A. Roy, *Nature* **362**, 56 (1993).
11. R. C. Thompson, T. P. Crowe, S. J. Hawkins, *Environ. Conserv.* **29**, 168 (2002).
12. K. R. Siegel, M. Kaur, A. C. Grigal, R. A. Metzler, G. H. Dickinson, *Ecol. Evol.* **12**, e8922 (2022).
13. B. Santelices, M. S. Doty, *Aquaculture* **78**, 95 (1989).
14. P. Rawiwan, Y. Peng, I. G. P. B. Paramayuda, S. Y. Quek, *Trends Food Sci. Technol.* **123**, 37 (2022).
15. J. Ollerton, *Pollinators & Pollination: Nature and Society* (Pelagic Publishing, 2021).

ACKNOWLEDGMENTS

J.O. and Z.-X.R. receive funding from the Chinese Academy of Sciences (2021VBA0014 and XDB31000000).

10.1126/science.add3198

MOLECULAR BIOLOGY

The material properties of chromatin in vivo

Chromatin is fluidlike within the crowded nucleus when probed in a living cell

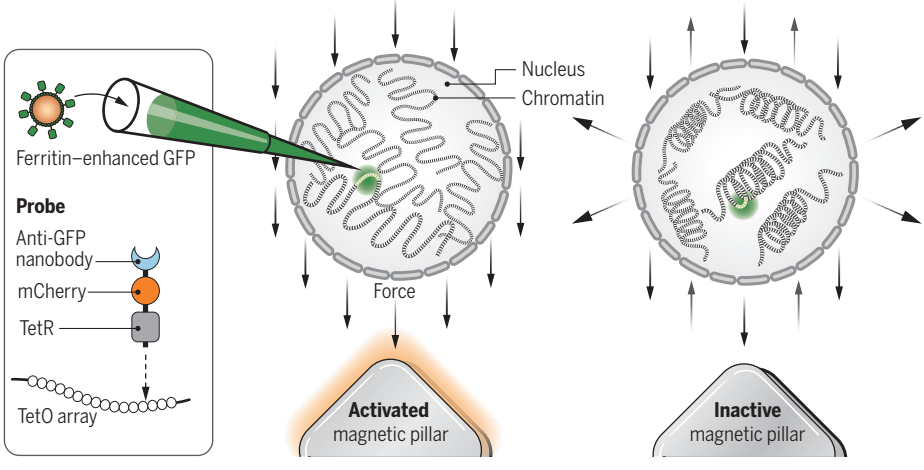
By **Woong Young So** *and* **Kandice Tanner**

The physical organization and material properties of the genome can regulate how genes are expressed and genome stability (*1*, *2*). On page 489 of this issue, Keizer *et al.* (*3*) report that chromatin behaves like a fluid in the nucleus of a living cell. Within minutes of applying magnetic forces directly to a marked locus within the genome of live cells, the authors observed displacement by several micrometers. Drawing on theoretical models in polymer physics, they determined that chromatin behaves as a free polymer in a viscous environment, with moderate interactions from the surrounding chromatin. These findings suggest that chromatin has a higher degree of flexibility than if it were to behave like a solid or gel-like material during interphase, as currently proposed (*4*–*7*). These results have implications for overall compaction of the genome and the coupling between mechanical and biochemical signaling needed for genome integrity and function.

Magnetic forces have previously been used to probe chromatin directly in living cells (*4*, *7*). However, these studies were unable to resolve single-chromatin dynamics. Keizer *et al.* solve the technical challenges of directly probing the mesoscale mechanical properties of chromatin in the context of local crowding in the surrounding nucleoplasm using magnetic nanoparticles as local sensors. Recombinant enhanced green fluorescent protein (eGFP)–labeled ferritin cages encapsulating a magnetic core were functionalized with a dimerizing protein construct of mCherry and anti-GFP fused with the tetracycline repressor (TetR). They were injected into the nuclei of cells that were engineered to contain numerous tetracycline operator (TetO) sequences within the genome. Upon docking with TetO, dimerization of the TetR facilitated binding to the locus (see the figure). These cells are cultured on magnetic micropillars that serve as conduits of subpiconewton forces after the application of an external magnetic field. These technical innovations supersede previous studies where the probes, a micrometer in diameter, were orders of magnitude larger than genomic loci (*4*). The application of forces

Probing chromatin behavior

Keizer *et al.* used an innovative probe to measure the mechanical properties of chromatin. A dimerizing protein construct of mCherry and anti–green fluorescent protein (GFP) nanobody fused with tetracycline repressor (TetR) binds to tetracycline operator (TetO) sequences within the genome. The addition of ferritin–enhanced GFP magnetic nanoparticles allows for movement of the chromatin with magnets. This reveals fluidlike behavior of chromatin (left) compared with unperturbed chromatin in the absence of the magnetic field (right).



GRAPHIC: N. CARY/*SCIENCE*

is also more physiologically appropriate than previously described; subpiconewton forces are associated with RNA polymerase II transcription and with motor proteins within the nucleus (*4*).

Chromatin must undergo conformational changes during and after replication, as observed during mitosis, rearrangements during meiosis, and structural folding as the cell cycle progresses (*7*). Conformational changes are also observed during nuclear deformations, as observed when immune and cancer cells traverse distinct tissues (*8*). Local chromatin dynamics influence how transcription factors dock to target sequences and regulate long-range interactions between distant sequences (*7*). The observation that chromatin is fluid prompts reexamination of the biophysical principles underpinning these processes.

Two discoveries reported by Keizer *et al.* require the revisiting of existing ideas. That subpiconewton forces are sufficient to physically drag a locus within the genome across several micrometers in a crowded nucleus directly contradicts a previous finding that chromatin remains confined with submicrometer displacements after applications of larger forces (*4*). Possible explanations to reconcile these findings are the differences in the size of the probe and the amount of time that the forces were applied. Keizer *et al.* applied weaker forces for several minutes, whereas the previous experiment involved probes that were ~1000-fold larger and applied higher forces for 100-fold shorter time scales (*4*). Also, the labeling technique in the current study relies on the insertion of an artificial locus, whereas a native locus might reveal different flexibilities. Advances in CRISPR-mediated editing could be used to endogenously label chromatin to reconcile these findings.

The results reported by Keizer *et al.* also challenge the view that chromatin behaves as a stiff, cross-linked polymer gel (*3*–*6*). Polymers, including chromatin, are often depicted as linearized assemblies of individual units covalently linked in a Gaussian chain (*9*). These chains can be modeled as beads connected with harmonic springs. Instead, the authors observed that chromatin has fluidlike properties and behaves as a free polymer. The Rouse model is usually used for concentrated polymer solutions, where a drag force can be used to understand forces on the polymer as it undergoes motion with minimal contributions resulting from hydrodynamics (*9*). In the study of Keizer *et al.*, the observed dynamics suggested that there are moderate contributions of hindrance, topology, and cross-linking because of interactions with other chromatin regions in a crowded nucleus. One explanation might be that there may be subdomains within chromatin, where a mixture of states of gel-like domains are interspersed within the larger liquid structure. An additional explanation might be that the strength of the cross-linking is weak as a result of short-lived cross-links within the network of chromatin chains and associated factors. Thus, heterogeneity in the material properties might exist at different length scales. This may further clarify the differential transcriptional activities observed for the two forms of chromatin found within the nucleus (euchromatin and heterochromatin) (*1*, *2*).

Phase separation has been shown to drive the formation of heterochromatin domains (*10*). The dynamics of integrated loci that are fluidlike may also be important for phase separation of chromatin in vivo. Further studies involving the manipulation of additional chromosomes using endogenous probes will address the generalizability of the findings. The study of Keizer *et al.* also opens possibilities to mechanically manipulate gene loci in vivo. The authors cultured cells on pillars. However, mechanotransduction of longer-range forces, such as shear stress, differential tissue rigidity, or osmotic stresses, influence local chromatin dynamics through direct cross-talk between the nucleus and the extracellular milieu (*11*). Thus, coupling these technical advances with culture models, such as three-dimensional organoids, can be used to address the knowledge gap between micro- and mesoscale behaviors. ■

REFERENCES AND NOTES

1. R. P. Martins, J. D. Finan, F. Guilak, D. A. Lee, *Annu. Rev. Biomed. Eng.* **14**, 431 (2012).
2. B. Fierz, M. G. Poirier, *Annu. Rev. Biophys.* **48**, 321 (2019).
3. V. I. P. Keizer *et al.*, *Science* **377**, 489 (2022).
4. A. H. B. de Vries, B. E. Krenn, R. van Driel, V. Subramaniam, J. S. Kanger, *Nano Lett.* **7**, 1424 (2007).
5. H. Strickfaden *et al.*, *Cell* **183**, 1772 (2020).
6. N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, C. Murre, *Nat. Commun.* **10**, 2771 (2019).
7. J. F. Marko, M. G. Poirier, *Biochem. Cell Biol.* **81**, 209 (2003).
8. Y. Kalukula, A. D. Stephens, J. Lammerding, S. Gabriele, *Nat. Rev. Mol. Cell Biol.* 10.1038/s41580-022-00480-z (2022).
9. P. E. Rouse Jr., *J. Chem. Phys.* **21**, 1272 (1953).
10. K. Rippe, *Cold Spring Harb. Perspect. Biol.* **14**, a040683 (2022).
11. Y. A. Miroshnikova, M. M. Nava, S. A. Wickström, *J. Cell Sci.* **130**, 2243 (2017).

ACKNOWLEDGMENTS

The authors are supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute.

10.1126/science.add5444

Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. Email: kandice.tanner@nih.gov

CARBON NANOTUBES

Engineering defects with DNA

Single-walled carbon nanotubes are structurally modified by using a genetic sequence

By **YuHuang Wang**

Strategically introduced defects can be used to modify carbon nanotubes for new properties and functions. For example, chemical defects can act as atomic traps for electrons, holes, electron-hole pairs, and even molecules and ions (*1*). The ability to control the placement of these defects on carbon nanotubes could enable a plethora of fundamental studies and potential applications in imaging, sensing, disease diagnostics, and quantum information science (*2*–*4*). However, there is yet no effective way to do this with atomic precision (*5*–*7*). On page 535 of this issue, Lin *et al.* (*8*) report the creation of ordered defect arrays by programming DNA strands to wrap around and traverse the entire length of a single-walled carbon nanotube.

Modifications of single-walled carbon nanotubes are typically random, as all reactive sites on the extensive hexagonal lattice of the carbon nanotube are effectively the same (*7*). Lin *et al.* show that it is possible to program the placement of defects on the carbon lattice by using single-stranded DNA. In one example, they use a specific DNA strand made of guanine (G) and cytosine (C), with the sequence CCCGCCCCCCCGCCC, to wrap around a single-walled carbon nanotube, with each of the two guanines covalently linked to the nanotube forming an atomic defect (see the figure). The two defect sites are separated by five carbon-carbon bonds along a helical line of the hexagonal lattice known as the "armchair line," which travels down the nanotube similar to the threads on a screw. The process created ordered arrays of defects and pinned multiple DNA strands along the armchair line.

Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. Email: yhw@umd.edu

Lin *et al.* observed this transformation using cryo–electron microscopy (cryo-EM), which measures the shadows of a macromolecule at many different orientations to reconstruct its three-dimensional (3D) structure. In their experiment, reconstructions from the cryo-EM images of 44,393 DNA-functionalized carbon nanotubes revealed that the single-stranded DNA forms a helical pitch of 6.5 Å, which closely matches the periodicity of the aforementioned armchair line of the carbon nanotube lattice. For comparison, this ordering is not observed in the controls. The formation of ordered defect arrays is also evidenced by spectral signatures that are strongly dependent on the spatial distribution pattern of the defects. When the two bonding G's are spaced at a distance that matches the helical pitch, the disorder-induced Raman scattering is minimized. Consistently, the defect photoluminescence peak shows minimum spectral shift and width at the same spacing between the G's and reveals a picture of populating electron-hole pairs, or excitons, between shallow defect traps in an array (*9*).

Although Lin *et al.* succeeded in designing a complex procedure for creating these defects, the exact mechanism behind the ordering remains somewhat of a mystery. The nanotubes are initially dispersed and stabilized individually in an aqueous solution by using a different DNA sequence (TTTCCCTTTCCCCCC; T is thymine), which is later displaced by bonding the G-containing sequence to the nanotube using 525-nm light to trigger the reaction. This guanine photochemistry creates covalent defects on single-walled carbon nanotubes, but the ordering of those defects was not previously observed (*6*). For these covalent defects to occur in an ordered array, the bonding reactions must occur in a way that avoids locking into a disordered configuration. To this end, the use of a water-methanol solvent might have allowed the defects to be created, as the methanol unwraps the dispersing DNA from the nanotube, making space for the reaction. Further research will be needed to decipher the mechanism of the ordering, which is important for designing and guiding the synthesis of these defect arrays.

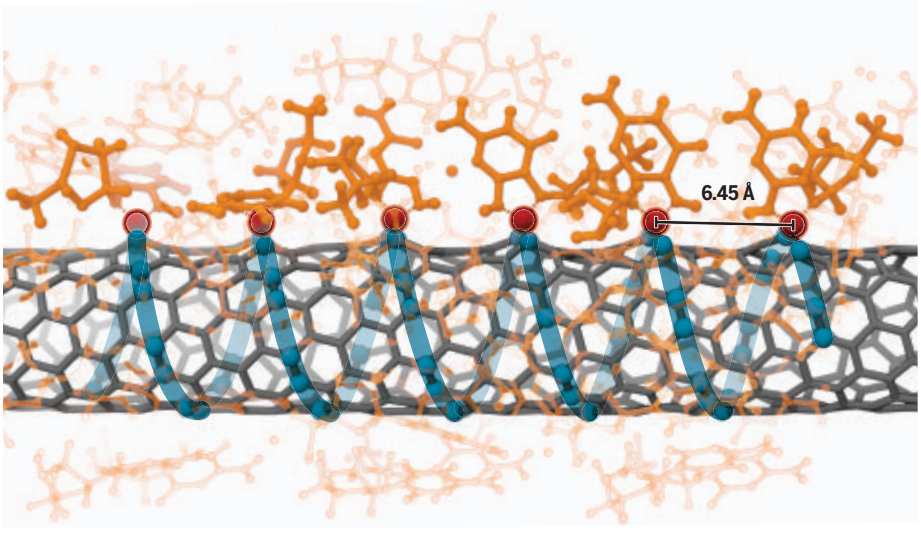
It is also worth highlighting the successful use of cryo-EM for resolving the oligonucleotides. Although cryo-EM has been powerful in determining the structure of proteins and other large biomolecules in the past, smaller molecules have typically been challenging for cryo-EM to resolve because of the low signal-to-noise ratio. The success here may be attributed, at least in part, to the nanotubes serving as an atomically precise ruler and reference for the 3D reconstruction. This "hack" may provide an exploit if it can be applied to resolving other small molecules that possess this property.

With a reliable method to create an ordered array of carbon nanotube defects, researchers may explore its applications—e.g., engineering defects as quantum traps to harness electron spins, excitons, trions, and single photons for a wide range of applications, such as in quantum chemical sensing (*1*). These traps, known as "sp^2" defects, are considerably shallower, and thus dimmer in photoluminescence, than "sp^3" defects, in which each carbon atom at the defect site is bonded to four instead of three other atoms (*10*). Although the weaker defect photoluminescence may limit the applications of these defects in areas such as imaging (*2*) and quantum light sources (*3*), their transport properties may be especially useful for a different assortment of applications. One such opportunity is for the experimental verification of the model for organic superconductors proposed more than 50 years ago by physicist Bill Little (*11*), in which he postulated that a room-temperature superconductor can be realized in a wire functionalized by ordered arrays of polarizable side chains that attract electrons. The finding by Lin *et al.* may provide a way to synthesize the wire imagined by Little and put the prediction to the test. The ability to program these defects along the length of a nanotube may also help create the so-called "Heisenberg spin chain," which can theoretically function as a chain of entangled quantum bits, with potential applications in quantum information processing and computing (*12*).

To realize these ambitious visions, future research efforts must expand and capitalize on the vast design space of DNAs, defect chemistry (*1, 13*), and precision nanotubes (*14*); enable deeper traps (*10*); and increase the size of the defect array. With the defect ordering demonstrated in this work, Lin *et al.*'s discovery (*8*) marks an important step, and the rush to order has started for defects. ■

Programming atomic defects on a carbon nanotube

By incorporating reactive guanines into a DNA sequence (orange), Lin *et al*. create ordered arrays of defects (red) at a pitch of ~6.5 Å along a helical line (blue) on the hexagonal lattice of a carbon nanotube (gray).



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. A. H. Brozena, M. Kim, L. R. Powell, Y. Wang, *Nat. Rev. Chem.* **3**, 375 (2019).
2. A. K. Mandal *et al.*, *Sci. Rep.* **10**, 5286 (2020).
3. X. He *et al.*, *Nat. Mater.* **17**, 663 (2018).
4. M. Kim *et al.*, *Nat. Biomed. Eng.* **6**, 267 (2022).
5. X. Wu, M. Kim, H. Qu, Y. Wang, *Nat. Commun.* **10**, 2672 (2019).
6. Y. Zheng, S. M. Bachilo, R. B. Weisman, *ACS Nano* **13**, 8222 (2019).
7. H. Qu *et al.*, *ACS Nano* **16**, 2077 (2022).
8. Z. Lin, *Science* **377**, 535 (2022).
9. M. Gurioli, A. Vinattieri, J. Martinez-Pastor, M. Colocci, *Phys. Rev. B* **50**, 11817 (1994).
10. Y. Piao *et al.*, *Nat. Chem.* **5**, 840 (2013).
11. W. A. Little, *Phys. Rev.* **134**, A1416 (1964).
12. H.-J. Mikeska, A. K. Kolezhuk, *Lect. Notes Phys.* **645**, 1 (2004).
13. T. Shiraki, Y. Miyauchi, K. Matsuda, N. Nakashima, *Acc. Chem. Res.* **53**, 1846 (2020).
14. F. Yang *et al.*, *Chem. Rev.* **120**, 2693 (2020).

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (grant nos. CHE2204202 and CHE1904488). I thank L. J. Wang for helping build the molecular model.

10.1126/science.abq2580

POLICY FORUM

GLOBAL HEALTH

A treaty to break the pandemic cycle

An evidence-based treaty must balance prevention, preparedness, response, and repair

By **Alexandra L. Phelan**[1,2,3] *and* **Colin J. Carlson**[1,2,4]

World Health Organization (WHO) member states have tasked an Intergovernmental Negotiating Body with developing an initial draft of a "pandemic treaty" by August 2022, with the goal of adoption of an agreement by May 2024. There are multiple proposals for the specific aims that such an international legal instrument for pandemic preparedness and response should incorporate. One unsatisfying solution would be to relegate the pandemic treaty solely to matters of human health as much as possible. International law, including the International Health Regulations (IHR), has typically operated in these types of fragmented, thematic silos. This untenable approach ignores key lessons from the past 2 years, and well before. A better option is a treaty that (pro)actively brings cohesion to international law and the haphazard governance that left the world vulnerable to both the risks and impacts of COVID-19.

Considerable scientific evidence suggests that the world is trapped in a positive-feedback loop: Disease spillovers from animals to humans become outbreaks that then become pandemics, creating many catastrophic impacts that can reduce resiliency and increase the socioecological drivers of spillovers. Pandemics connect to every aspect of society, both driven by and spilling over into politics, the environment, economics, and systematic inequalities and injustices. This imposes a substantial burden on the negotiation process, as stakeholders lobby for the treaty to be all things for all interests. Left unchecked, this might result in a treaty that fails at its primary purpose: preventing, preparing for, and responding to pandemics.

[1]Department of Microbiology and Immunology, Georgetown University Medical Center, Washington, DC, USA. [2]Center for Global Health Science and Security, Georgetown University Medical Center, Washington, DC, USA. [3]O'Neill Institute for National and Global Health Law, Georgetown University, Washington, DC, USA. [4]Department of Biology, Georgetown University, Washington, DC, USA.
Email: alexandra.phelan@georgetown.edu

Closing the "pandemic gap" requires evidence-based solutions that break every step of the cycle, including reducing spillover risk, reducing pandemic risk, reducing pandemic impacts, and ensuring recovery and resilience. We identify 12 elements for a cohesive, transformative, and evidence-based treaty (see the box). We describe how these elements could be translated into provisions under the treaty, with a focus on principles of good governance, distributive justice, and human rights and coherence within international law. A treaty that balances these aims will move global health governance beyond the limited scope of the IHR and provide clarity and complementarity to other relevant international legal regimes.

SPILLOVER PREVENTION

Pandemics start in nature. Stopping some at the source could be cost effective, and many proposed solutions would incidentally benefit conservation and other environmental causes.

Planetary health solutions

Anthropogenic environmental change drives poorer baseline human and animal health and is increasing the rate of viral spillover. To date, land use has been the biggest upstream driver, linked to 18% of emerging disease outbreaks since 1940 (*1*). Within the next decade, it will be overtaken by climate change, even if greenhouse gas emissions are reduced (*2*). Global environmental issues must be embedded throughout the treaty. This includes preambular language that expressly recognizes the roles that climate change, biodiversity loss, land degradation, and wildlife trade play in pandemic risk. Other elements of the treaty may also support environmental aims: For example, obligations to build sustainable primary health care systems may help reduce deforestation and meet greenhouse gas emissions targets (*3*). To formalize cobenefits, a treaty can expressly interface with existing obligations under other international laws, such as the Convention on Biological Diversity (CBD), the Convention on International Trade in Endangered Species (CITES), and the United Nations Framework Convention on Climate Change (UNFCCC).

One Health solutions

Outbreaks are multicausal: A recent analysis of 41 socioecological risk factors (such as mining, hunting, and travel) showed that in the 100 largest zoonotic disease outbreaks since 1975, no one driver could be linked to more than 13% of events (*4*). Singular solutions that only target one aspect of the wildlife-livestock-human interface are therefore unlikely to succeed at scale. Taking lessons from multilateral environmental agreements, and in coherence with states' existing obligations in relation to the World Organisation for Animal Health (WOAH, formerly OIE) and the Food and Agricultural Organization (FAO), the treaty should provide flexibility for locally tailored risk assessments and intervention strategies while instilling obligations to conduct animal disease surveillance and fund a stable, well trained One Health workforce (integrating human, animal, and environmental health). Historically, these investments have been focused in tropical countries and around high-risk interfaces like wildlife markets and agriculture; a forward-looking pandemic treaty must expand this approach, cognizant of scientific evidence that, particularly in a changing climate, zoonotic emergence is a global risk (*2*).

Zoonotic risk assessment

After facing five influenza pandemics in less than a century, global strategies have been established to sound the alarm when new threats emerge. COVID-19 underscores the need to broaden this approach: Although flu is the most persistent pandemic threat, there are hundreds of coronaviruses, paramyxoviruses, poxviruses, flaviviruses, and others that will cross the species divide in the coming century. Only a small fraction of spillovers are likely to become pandemics; the task of risk assessment therefore relies on both surveillance (including viral discovery in animals, syndromic surveillance in humans, and monitoring at the interface between them) and science (including

both experimental characterization and computational prediction). Establishing notification and data-sharing obligations for animal disease outbreaks and One Health discoveries of public interest, coupled with both risk assessment and benefit sharing provisions, would support global efforts to identify potential zoonoses, develop countermeasures like universal vaccines, and monitor for spillover.

PANDEMIC PREVENTION

Even with upstream mitigation, spillover remains a persistent problem. The primary aim of a pandemic treaty must be to prevent outbreaks from becoming pandemics.

Surveillance and assessment

Infectious disease surveillance is severely limited by funding, workforce, and technological capacity. Genomic epidemiology is vital both to identify outbreaks and to assess the risk of human-to-human spread (*5*). Under the IHR, there is no express obligation to share pathogen samples or genomic sequence data (GSD), even after notification of a potential public health emergency of international concern. The pandemic treaty must consider the balance of obligations to share data, obligations on high-income countries to provide capacity building and technology transfer, and obligations to equitably share the benefits from the use of GSD, including not only risk assessments but also diagnostics, vaccines, and treatments. A treaty may also support innovation in disease detection (including wastewater surveillance for both new and ongoing outbreaks) and real-time digital sharing of epidemiological data.

Biomedical R&D and production

Under one estimate, $24 billion would be sufficient to develop prototype vaccines for every family of viruses that can infect humans (*6*). With biomedical investments like these, and international initiatives like the Coalition for Epidemic Preparedness Innovations, scientific advances could lead to game-changing vaccine platforms and technologies, potentially including universal vaccines. For production and deployment to keep pace with these technological advances, the pandemic treaty must address the substantial legal and political impediments to technology transfer and capacity building. This includes taking lessons from other regimes, including international climate law, for financing and technology mechanisms to build capacities and facilitate hard and soft technology transfer. A treaty can set priorities and establish an end-to-end research and development (R&D) platform, predictable financing, and regional platforms and engage industry and states in prenegotiated contracts and agreements from R&D to production (*7*).

Health systems strengthening

Even before vaccination began, health care disparities made COVID-19 twice as deadly in low- and middle-income countries (*8*). The goal of achieving both universal health coverage (UHC) and global health security must form the foundation of a pandemic treaty if it aims to prevent outbreaks from becoming pandemics. The treaty must explicitly acknowledge the importance of health systems strengthening (HSS) in achieving both UHC and global health security and use a "right to health" approach to ensure that health systems are accessible, available, acceptable, and of sufficient quality to all. This should include express commitments to HSS and UHC, including financial commitments: Countries' investment choices will account for an 870-million-person margin of uncertainty in projected health coverage by 2030 (*9*). To achieve the best outcomes possible, a treaty must include a funding mechanism with a target of meeting the WHO's existing goal of 1 billion more people covered by UHC. This would not only be fundamental to prevent spillovers from becoming pandemics but would bring a suite of health cobenefits, like reducing maternal and child mortality and the burden of neglected diseases, as well as reducing pandemic impacts.

Twelve elements of a pandemic treaty

A successful treaty must balance several aims between prevention (reducing both spillover and pandemic risk) and response (reducing pandemic impacts and improving recovery and resilience).

Reduce spillover risk
- Planetary health solutions (global)
- One Health solutions (local)
- Zoonotic risk assessment

Reduce pandemic risk
- Surveillance and assessment
- Biomedical R&D and production
- Health systems strengthening

Reduce pandemic impacts
- Equitable access to global goods
- Emergency legal preparedness
- Least restrictive measures

Recovery and resilience
- Adaptive governance
- Accountability and transparency
- Reduce inequities and injustice

PANDEMIC RESPONSE

As of now, there is no international law that sufficiently governs a pandemic response, leading to the kinds of dysfunction witnessed during COVID-19.

Equitable access to global goods

The treaty must address the global injustice of the inequitable sharing of benefits from the use of pathogen samples and GSD, including vaccines, diagnostics, and therapeutics seen in current and previous pandemics. The treaty is the most critical opportunity that we have to establish a multilateral mechanism to facilitate this in line with global health and equity and consistent with other areas of international law, like the Nagoya Protocol to the CBD. In addition, the treaty should limit legal barriers to global equitable access, such as export controls, and set the terms of use for advance purchase agreements, including the use of prenegotiated standard terms and establishment of a multilateral distribution mechanism (*7*). The treaty is an opportunity to dismantle intellectual property law barriers to equitable access in a forum that is more democratically representative of the Global South than the World Trade Organization, potentially better enabling the prioritization of global health and equity.

Emergency legal preparedness

Law has long been undervalued as a critical component of emergency preparedness and response. However, legislative and judicial actions throughout the pandemic have left many countries' public health laws in tatters (*10*). Governments are substantially less empowered to protect health now and during the next pandemic. Although the IHR require States Parties to have enabling legislation as part of their core public health capacities, the contents of these obligations are poorly defined across implementation initiatives. It was only this year that the Global Health Security Agenda, aimed at strengthening global health security by building core public health capacities worldwide, added a program sufficiently dedicated to legal preparedness. Translating the IHR's technical approaches into domestic law has been a major challenge, partly because domestic laws depend on political will for action. Here, a pandemic treaty that is not designed exclusively as a technical instrument but also carries normative weight may be better equipped to facilitate the political momentum necessary for domestic law reform.

Least restrictive measures

Public health and human rights demand that governments use the least restrictive measures necessary to achieve proportionate public health outcomes. The pandemic treaty must provide the governance necessary to guide evidence-based public health measures and respect for human rights. It must also establish mechanisms to monitor restrictive public health measures used during pandemics and ensure accountability of states under global health law (*11*). This includes addressing the use of travel restrictions. Scientific evidence indicates that international travel restrictions may delay the spread of pandemic pathogens (*12*), but they have been applied inconsistently and discriminatorily and some have been more restrictive than necessary. The use of travel restrictions has serious implications for the next pandemic, potentially disincentivizing rapid notifications of new outbreaks. In coherence with the IHR and any reforms, the pandemic treaty is an opportunity to reconsider the types of incentives available for rapid reporting, including funding and compliance mechanisms, as well as for normative commitments to least restrictive measures at the highest levels of government.

RECOVERY AND RESILIENCE

A pandemic treaty offers the best, and possibly last, chance to rebuild global trust and show that multilateralism has the political and practical ability to build a safer and fairer world.

Adaptive governance

Responding to the changing landscape of global health will require a treaty to establish a regular conference of parties (COP). These meetings will facilitate norm-building, interpretation consensus, compliance assessment, and the negotiation of protocols if a framework convention model is adopted. Learning from the relationship between the UNFCCC and Intergovernmental Panel on Climate Change (IPCC) [and, similarly modeled on it, the CBD and the Intergovernmental Panel on Biodiversity and Ecosystem Services (IPBES)], the treaty should also establish an intergovernmental science-policy platform for pandemic prevention, preparedness, and response, which will inform decision-making at COPs. This platform must have a primary mandate of evidence synthesis across all four quadrants of the pandemic cycle, explicitly not in support of any agenda beyond consolidation and evaluation of scientific literature, with an independent public process for government and civil society review. Building on this, and taking the lessons from the Biological Weapons Convention, the treaty should build in express opportunities for regular review conferences, ensuring that the treaty itself survives and adapts to changing science, technologies, and epidemiology.

Accountability and transparency

The COVID-19 pandemic has decreased public trust in both scientists and government officials. Treaty elements that promote accountability and transparency will be essential to its success and good governance, establishing the rule of law for state conduct. Similar to mechanisms under international climate law, the treaty should include an independent compliance body tasked with facilitating implementation and assistance, conducting verification, and promoting compliance through transparent and regular reporting (*13*). A range of other compliance mechanisms have been proposed, from empowering WHO or the United Nations Secretary-General to conduct investigations to sanctions for delayed notifications. However, punitive measures risk undermining the trust and transparency necessary for public health, even at the global scale. Instead, the treaty should establish transparency and accountability through positive incentives. Civil society will also be a critical component of holding states accountable for their obligations under the treaty, including in any shadow reporting processes, and protecting the rights of health care workers, scientists, and lawyers who serve as whistleblowers (*14*). In recent years, many governments have actively limited the space for civil society to operate, share information, and conduct advocacy in public health; to remedy this, the treaty must enshrine recognition of civil society organizations and factor in appropriate protections and participation in compliance processes.

Reduce inequities and injustice

Throughout all of the elements, there must be an express commitment by States Parties to reduce the inequities and injustices that make pandemics disproportionately affect vulnerable populations, both within and among countries. Although this recognition should include preambular language that sets the tone for the treaty, it must go further. The treaty must contain a clear, binding obligation of nondiscrimination in the implementation of pandemic prevention, preparedness, and response measures. The negotiation process, the science-policy platform, and the treaty's governance arrangements must all have express requirements for equitable geographical representation and diversity of gender, country income levels, and disciplinarity. Finally, the treaty must not be given the opportunity to recapitulate the colonial histories of global health and international law. Deliberations about how a pandemic treaty enshrines and advances the decolonization of global health should be central to its conception and led by countries and communities that have historically been excluded from both global governance and scientific scholarship.

BREAKING THE CYCLE

Breaking the pandemic cycle and dismantling injustice are daunting tasks for a treaty, particularly one designed within a system that has perpetuated and institutionalized both. Even this early in negotiations, many already anticipate that on the current trajectory, a treaty will fall short of these ambitions (*15*). Negotiators must take these concerns as a clarion call or risk far-reaching repercussions for the legitimacy of global health and multilateralism beyond the treaty itself. Every remaining moment is a once-in-a-generation opportunity to negotiate an ambitious and equitable treaty and fight for provisions that materially improve the health of humans, animals, and ecosystems. ■

REFERENCES AND NOTES

1. F. Keesing *et al.*, *Nature* **468**, 647 (2010).
2. C. J. Carlson *et al.*, *Nature* 10.1038/s41586-022-04788-w (2022).
3. I. J. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28515 (2020).
4. P. R. Stephens *et al.*, *Philos. Trans. R. Soc. London Ser. B* **376**, 20200535 (2021).
5. National Academies of Sciences, Engineering, and Medicine *et al.*, *Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies* (National Academies Press, 2020).
6. N. Teran, "Why BARDA deserves more funding" (Institute for Progress, 2022); https://progress.institute/why-barda-deserves-more-funding/.
7. R. Ramchandani *et al.*, *BMJ* **375**, e067488 (2021).
8. A. T. Levin *et al.*, *BMJ Glob. Health* **7**, e008477 (2022).
9. Global Burden of Disease Health Financing Collaborator Network, *Lancet* **391**, 1783 (2018).
10. The Network for Public Health Law, National Association of County and City Health Officials (NACCHO), *Proposed Limits on Public Health Authority: Dangerous for Public Health* (The Network for Public Health Law, NACCHO, 2021).
11. T. F. Hodgson *et al.*, "Human rights must guide a pandemic treaty" (Health and Human Rights Journal, 2021); https://www.hhrjournal.org/2021/11/human-rights-must-guide-a-pandemic-treaty/.
12. L. Besançon, A. Flahault, G. Meyerowitz-Katz, *J. Travel Med.* **29**, taac055 (2022).
13. A. L. Phelan, *Nature* **605**, 397 (2022).
14. S. E. Davies, *Med. Law Rev.* **25**, 270 (2017).
15. C. Wenham, M. Eccleston-Turner, M. Voss, *Int. Aff.* **98**, 837 (2022).

ACKNOWLEDGMENTS

This work was supported by the Carnegie Corporation of New York.

Published online 14 July 2022
10.1126/science.abq5917

Taxi drivers in Shenzhen, China wait for their electric vehicles to charge.

BOOKS *et al.*

BATTERIES

The hidden costs of batteries

Clean energy plans must include safer, more sustainable, worker-friendly power storage

By **Benjamin K. Sovacool**

Batteries are everywhere—in our mobile phones, our remote controls, our toys, and increasingly in the fleets of electric vehicles on our roads. Batteries are ubiquitous because they can do many things: They can store energy in homes, improve the resilience of electric grids, and assist with the integration of low-carbon electricity sources such as wind farms and solar photovoltaic panels.

Batteries, argues environmental historian James Morton Turner in his new book, *Charged*, are the cornerstone of our modern "culture of mobility," one that depends on portable, reliable, and cordless sources of energy. In *Charged*, Turner offers an eminently readable, elegantly precise treatise on the topic of batteries. The book consists of only four main chapters. An introduction describes the role of batteries in our modern economy, and a conclusion outlines recommendations for cleaner forms of energy. In between, the book includes a chapter on lead-acid batteries and cars, a chapter on AA batteries and consumer or "throwaway" goods, a chapter on lithium-ion batteries and smart phones, and a chapter on electric mobility. The book even offers tips on how to make a homemade battery from a lemon.

Charged is more critical of battery power than most other books on this topic, noting that they appear "clean" only because they obscure other, more discrete, impacts. In the environmental dimension, batteries depend on some of the most toxic metals and materials in industrial supply chains. The list of environmental insults is long and sobering: lead, sulfuric acid, mercury, manganese, zinc, steel, carbon, graphite, ammonium chloride, potassium hydroxide, cadmium, lithium, nickel, cobalt, and sometimes rare earth minerals. Batteries are thus intimately connected with mining and extraction and with energy-intensive processing. When they are used up, they often return to the earth in the hazardous form of waste or pollution.

The material economy behind batteries also has a poignant human and social cost. People at both ends of the battery supply chain—extraction and waste—must handle the toxic materials they contain, either coping with the precarious practices of artisanal mining for lithium and cobalt or managing burgeoning flows of electronic waste at scrapyards. Every step in the battery manufacturing process moves toxic materials such as lead or cobalt into workers' bodies, notes Turner, "most often through what they ate, drank, or, most problematically, breathed."

These issues provoke Turner to recommend some fairly substantial, and likely controversial, recommendations. He believes that governments must lead to build a clean energy future from the ground up. He suggests that policies are needed to promote more responsible sourcing of battery materials and more responsible manufacturing and production of batteries. He argues in favor of new and expanded mining operations (as well as more sustainable refining operations) in the United States. Most provocatively, he argues that planners should cease "trying to recycle our way to a clean energy future" and reject "circular economy" ideas in favor of more practical materials and supply chain management within the battery production process, including more responsible mining practices.

In elucidating the complex history of batteries in our society, the book offers a much needed exposé on the human and environmental side of proposed low-carbon transitions that depend on batteries or energy storage. The book also underscores how energy choices are always about more than energy technology or resources. They intertwine with, and even accelerate, practices of automobility, hyperconsumption, digitalization, and social media. Moreover, the book demonstrates that energy security plans and energy policymaking are not, at their core, adequately addressing the many social and environmental dilemmas introduced by batteries.

Charged
James Morton Turner
University of Washington Press, 2022. 256 pp.

The risk, as Turner states at the end of the book, is immense: "scaling up a clean energy future could very easily drive social inequities and concentrate economic power, just as the fossil fuel–powered extractive economy of the twentieth century did." But so is the ultimate reward for designing safer, more sustainable, worker-friendly batteries, which could become the foundation for our next global energy transition. ■

10.1126/science.add0714

The reviewer is at the Department of Earth and Environment, Boston University, Boston, MA, USA; the Science Policy Research Unit, University of Sussex, Brighton, UK; and the Department of Business and Technology Development, Aarhus University, Herning, Denmark. Email: b.sovacool@sussex.ac.uk

PHOTO: BILLY H. C. HWOK

BIOTECHNOLOGY

The ascent of Moderna

From scrappy startup to household name, a journalist traces the trajectory of the company and its mRNA vaccine

By **Elie Dolgin**

In early 2020, US government scientists were preparing to conduct a drill to determine how rapidly a new vaccine technology could be deployed to combat an emerging pathogen. The deadly Nipah virus had recently leapt from bats into people in southern India, killing nearly everyone who was infected. The plan: to see how quickly researchers at the National Institutes of Health, in collaboration with the biotech company Moderna, could build a vaccine from scratch and launch a clinical trial, all using the then-unproven messenger RNA (mRNA) vaccine platform.

The drill, however, soon took a backseat to a real respiratory virus that had begun sickening dozens in central China. In a record-setting 66 days, Moderna and its government partners went from having a genetic sequence of the new coronavirus to a first-in-human vaccine trial. A US Food and Drug Administration–authorized vaccine followed just 10 months later, and Moderna became a household name seemingly overnight.

Peter Loftus, a reporter for *The Wall Street Journal*, was quick to recognize the historic magnitude of Moderna's vaccine project in its earliest days. He visited the company's headquarters in Cambridge, Massachusetts, in mid-February 2020, weeks before any cases of community transmission had even been documented on U.S. soil. He broke news of the shipment of Moderna's first batch of clinic-bound vaccine later that month and stayed in touch with company executives. His new book, *The Messenger*, offers readers an insider's view of the company's ascent from secretive startup to one of the most valuable health care companies in the world.

The protagonist of *The Messenger* is longtime Moderna chief executive Stéphane Bancel. An ambitious businessman with an unwavering belief in his company's technology, Bancel is, as Loftus writes, a "wizard at fundraising."

Before the pandemic, Bancel had amassed staggering sums of money on the promise that mRNA could be turned into powerful treatments for genetic diseases, cancer, heart failure, and more. When that idea did not pan out, the company pivoted to vaccines for infectious diseases and used data from a small, proof-of-concept flu vaccine trial to bring in more cash. With each new infusion of capital came hiring sprees and infrastructure investments, along with pilot projects for other viruses with pandemic potential, all of which meant that Moderna had a vaccine platform ready when COVID-19 came along.

Loftus captures Bancel's idiosyncrasies with some amusing personal details. He reveals, for example, how the Moderna CEO once gifted a Hermès belt to an academic who had previously teased him about the buckle design. The journalist's eye is drawn to the many paperweights that adorn Bancel's bookshelf—each containing a tiny vial of drug or vaccine encased in glass.

The book also discusses Bancel's brusque management style that rubbed many employees the wrong way. He yelled. He demeaned. "He hired people and made their life hell," as one former scientist tells Loftus. However, despite noting unusually high rates of staff turnover and previously chronicled complaints of a toxic work environment, the book does not offer many new fly-on-the-wall details on this topic.

Instead, Loftus focuses on money—how much Moderna raised from investors, the fluctuations of its stock price, the compensation packages of its executives, the billionaire status of its founders, the taxpayer-backed contracts that underwrote some of the company's riskiest bets. The company, we learn, raked in more funding faster than just about any other in biotech history, giving it advantages of scale and prestige.

But Moderna did not invent the idea of mRNA-based medicines. It was not the first to win authorization for an mRNA-based COVID-19 vaccine. And yet, in Loftus's telling, "It's largely Moderna that is being credited with establishing the disruptive mRNA technology."

"A doubted underdog driven by relentless idealists," Moderna is presented as the biotech David against pharma Goliaths. It is true that an industry giant (Pfizer) helped BioNTech—the German company that beat Moderna to the mRNA vaccine punch—get its COVID-19 shot over the finish line. But Moderna was not operating unaided—it had the weight of the US government behind it.

Moderna deserves a great deal of credit for its decade-long slog to push mRNA medicines forward. But just as with the formulation of calculus in the 17th century, the discovery of oxygen in the 18th century, or the theory of evolution in the 19th century, mRNA vaccine technologies were in the 21st century ether, ready for development.

Now, flush with profits from its COVID-19 vaccine, Moderna needs to prove that it is more than a one-trick pony. All eyes are on the messenger to see if it will be the bearer of good vaccine news to come. ■

10.1126/science.add1577

The reviewer is a science journalist in Somerville, MA, USA. Email: elie@eliedolgin.com

The Messenger: Moderna, the Vaccine, and the Business Gamble That Changed the World
Peter Loftus
Harvard Business Review Press, 2022. 320 pp.

SCIENCE & FOOD

PHOTO: MARKO GEBER/GETTY IMAGES

Campbell recommends a primarily plant-based diet.

PODCAST

The Future of Nutrition: An Insider's Look at the Science, Why We Keep Getting It Wrong, and How to Start Getting It Right
T. Colin Campbell with Nelson Disla
BenBella Books, 2020. 336 pp.

Having spent his youth on a dairy farm, biochemist T. Colin Campbell was an advocate for the consumption of meat until his own research began to suggest that diets high in animal protein may increase the risk of developing certain diseases. This week on the *Science* podcast, Campbell discusses how food can influence health and details the benefits of consuming a plant-focused diet. https://bit.ly/3yO1cOO

10.1126/science.add8277

LETTERS

Edited by **Jennifer Sills**

Funding databases for Ukrainian academics

The ongoing Russian invasion jeopardizes research in Ukraine. More than 200 Ukrainian educational institutions have been destroyed (*1*), and an estimated several thousand academics have fled the country (*2*, *3*). However, the majority of researchers—about 70,000 people—remain in Ukraine, where they are struggling to make a living and continue their research (*2*, *3*).

The international science community has provided generous support to Ukrainian researchers. Fellowships have been made available to displaced academics (*4*, *5*). Universities have offered remote working opportunities (*6*). National funders have stepped up with open calls for Ukrainian academics (*7*, *8*), and displaced-academics programs have prioritized Ukrainians (*9*).

Initially, most job and funding offers were advertised locally on university or funding body websites as well as on Twitter. However, many offers have gone unnoticed. To find them, Ukrainians must navigate the academic systems of foreign countries and search multiple websites, many of which list offers only in the country's native language. Moreover, the Ukrainian research community does not use Twitter regularly, so they are unlikely to see a poorly promoted tweet (*3*).

Several organizations have addressed these challenges by creating centralized databases that collect funding opportunities for Ukrainian researchers. Some databases are specific to a discipline, such as the one curated by the European Molecular Biology Organization (*10*). The European Commission has funded the Euraxess database, covering opportunities across Europe (*11*). The #ScienceForUkraine database, managed by volunteers, collects worldwide opportunities for Ukrainians (*12*).

We encourage institutions and funders who would like to support the Ukrainian academic community to streamline the process by including their offers in established databases. When advertising on social media, institutions can make offers easier to find by tagging them with widely known hashtags such as #ERA4Ukraine and #ScienceForUkraine. Ukrainians trying to continue their research in wartime need a straightforward way to access support.

Valentina Mosienko[1]*, Marina Pelepets[2], Sanita Reinsone[3], Michael Rose[4]
[1]School of Physiology, Pharmacology, and Neuroscience, University of Bristol, Bristol, UK. [2]Munich, Germany. [3]Institute of Literature, Folklore, and Art, University of Latvia, Riga, Latvia. [4]Max Planck Institute for Innovation and Competition, Munich, Germany.
*Corresponding author.
Email: valentina.mosienko@bristol.ac.uk

REFERENCES AND NOTES

1. Ministry of Education and Science of Ukraine, "Overview of the current state of education and science in Ukraine in terms of Russian aggression (as of June 13–25, 2022)" (2022); https://reliefweb.int/report/ukraine/overview-current-state-education-and-science-ukraine-terms-russian-aggression-june-13-25-2022.
2. G. Naujokaitytė, "Ukrainian researchers want to shift the focus of international support from crisis response mode to rebuilding the science ecosystem," *ScienceBusiness* (2022).
3. M. E. Rose *et al.*, "#ScienceForUkraine: An initiative to support the Ukrainian academic community: '3 Months Since Russia's Invasion in Ukraine,' February 26 – May 31, 2022," Max Planck Institute for Innovation & Competition Research Paper No. 22-13 (2022).
4. Volkswagen Foundation, Funding for Refugee Scholars and Scientists from Ukraine (2022); www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/funding-for-refugee-scholars-and-scientists-from-ukraine.
5. Polish Academy of Sciences, Joint Support for Ukrainian Scholars (2022); https://institution.pan.pl/index.php/769-support-for-ukrainian-scholars.
6. Indiana University, "IU Bloomington to host nonresidential scholars from Ukraine" (2022); https://news.iu.edu/stories/2022/06/iub/releases/06-host-nonresidential-scholars-from-ukraine.html.
7. Study in Latvia, "Support for UA students" (2022); https://studyinlatvia.lv/ua/scholarships-2.
8. Swiss National Science Foundation, "Support measures for scientists from Ukraine" (2022); www.snf.ch/en/WpOfVoVBYdBid6By/news/support-measures-for-scientists-from-ukraine.
9. CARA: Council for At-Risk Academics, Funding Directory—Ukraine (2022); www.cara.ngo/funding-directory-ukraine.
10. European Molecular Biology Organization, Solidarity with Ukraine and its life scientists (2022); www.embo.org/solidarity-with-ukraine.
11. European Commission, Euraxess database, "ERA for Ukraine" (2022); https://euraxess.ec.europa.eu/ukraine/offers.
12. #ScienceForUkraine (2022); https://scienceforukraine.eu.

COMPETING INTERESTS
S.R. and M.R. are founders of #ScienceForUkraine, M.P. is a coordinator of the German branch of #ScienceForUkraine, and V.M. is a volunteer for a UK branch of #ScienceForUkraine.

10.1126/science.add9155

A Russian rocket partially destroyed this educational and laboratory building at a college in Kharkiv in June.

PHOTO: SERGEY BOBOK/GETTY IMAGES

Chinchilla conservation vs. gold mining in Chile

Since pre-Columbian times, both gold and chinchilla fur have been in demand (*1*). From the first conquistadors to modern times, these resources have been severely exploited in the South American Andes, degrading the environment (*2*, *3*) and driving chinchillas to the brink of extinction (*4*). Now, gold mining is threatening some of the last wild chinchilla populations in Chile (*5*). Fortunately, Chile's environmental laws (*6*) have so far worked as designed to protect this Endangered species (*7*).

Before the 19th century, the two species of chinchilla—long-tailed or coastal (*Chinchilla lanigera*) and short-tailed or Andean (*Chinchilla chinchilla*) (*8*)—spanned the coast and the Andes regions of Chile, Peru, Bolivia, and Argentina. As of the mid-20th century, chinchillas were thought to be extinct from their native ranges, but recently a few scattered colonies of short-tailed chinchillas were discovered in some of the most inhospitable and remote places of Chile and Bolivia (*9*, *10*). These animals, likely recovering from intense human persecution over the past 200 years, need protection from human activities to thrive.

However, large amounts of gold have been discovered beneath the chinchilla colonies. The Gold Fields mining sites in northern Chile (*5*, *11*), for example, hold ~3.5 million ounces of extractable gold. The mining company is ready to begin extraction, but Chilean law prohibits

development projects from harming local biodiversity (*6*). The mining company has proposed a solution of moving the chinchillas, but its attempts to translocate a small number of chinchillas failed (*12*). Meanwhile, paid lobbyists are attempting to ease the restrictions to allow gold extraction to move forward.

Short-tailed chinchillas live in one of the most extreme environments known for a mammal: rocky outcrops, barren except for scattered grasses, located at elevations of 4000 m or higher. Challenges to population growth include slow reproduction and small, dispersed social groups (*4*, *10*). Mining poses a grave threat to the fragile equilibrium of these populations. The Chilean government should continue to abide by the current laws, which require the viability of translocations to be thoroughly evaluated and tested before the chinchilla habitat is destroyed by mining, both in this case and in future conflicts between corporate goals and environmental conservation. Other countries, including Bolivia, should also use this framework as a model to protect vulnerable species within their borders.

J. E. Jiménez[1]*, A. Deane[2], L. F. Pacheco[3], E. F. Pavez[4], J. Salazar-Bravo[5], P. Valladares Faúndez[6]
[1]Department of Biological Sciences and Advanced Environmental Research Institute, University of North Texas, Denton, TX 76201, USA. [2]Small Mammal Specialist Group Chinchilla Species Coordinator, Gainesville, FL 32605, USA. [3]Colección Boliviana de Fauna, Instituto de Ecología, Carrera de Biología, Universidad Mayor de San Andrés, La Paz, Bolivia. [4]Bioamérica Consultores, Santiago, Chile. [5]Department of Biological Sciences and Director of the International Center for Arid and Semiarid Land Studies, Texas Tech University, Lubbock, TX 79409, USA. [6]Department of Biology, Sciences Faculty, University of Tarapacá, Arica, Chile.
*Corresponding author: jaime.jimenez@unt.edu

REFERENCES AND NOTES

1. J. M. Maryanski, A. E. Nielsen, *Archaeofauna* **24**, 225 (2015) [in Spanish].
2. S. A. Abdul-Wahab, F. A. Marikar, *Central. Eur. J. Eng.* **2**, 304 (2012).
3. N. L. Alvarez-Berríos, T. M. Aide, *Env. Res. Lett.* **10**, 014006 (2015).
4. J. E. Jiménez, *Biol. Conserv.* **77**, 1 (1996).
5. E. Stoddard, "The chinchillas and the gold mine," *Undark* (2020).
6. Ministerio Secretaría General de la Presidencia, Chile, "Ley 19300, sobre bases generales del medio ambiente" (2020) [in Spanish].
7. "Short-tailed Chinchilla" (IUCN Red List of Threatened Species, 2016); www.iucnredlist.org/species/4651/22191157.
8. A. E. Spotorno, J. P. Valladares, J. C. Marin, R. E. Palma, C. Zuleta, *J. Mamm.* **85**, 384 (2004).
9. P. Valladares, A. E. Spotorno, A. Cortes, C. Zuleta, *Mammal. Spec.* **50**, 51 (2018).
10. E. Delgado, L. F. Pacheco, J. Salazar-Bravo, O. Rocha, *Oryx* 52, 13 (2018).
11. E. Stoddard, "A high-stakes chinchilla relocation effort stalls," *Undark* (2021).
12. Superintendencia del Medio Ambiente, Gobierno de Chile, "Resolución exenta N°1/Rol D-246-2021: Formula cargos que indica a minera Gold Fields Salares Norte, titular del proyecto Estudio de Impacto Ambiental Salares Norte" (2021); https://snifa.sma.gob.cl/Sancionatorio/Ficha/2757 [in Spanish].

10.1126/science.add7709

PHOTO: M.E. GARCI/IIM-CSIC

An octopus approaches the authors' camera.

LIFE IN SCIENCE

A scuffle at sea

Submerged 15 meters deep beneath the waters of Spain's Atlantic Islands National Park, we were on a mission to identify cephalopod spawning habitats, information that would inform the design of a new Marine Protected Area (MPA). We had been scuba diving in these cold waters for several days, recording data on octopuses, cuttlefish, and squid. Visibility was good, and we looked forward to what promised to be a successful dive.

In the distance, we spotted a large octopus that was camouflaged on the sandy ocean floor. We carefully approached it to take some photos, but the octopus soon realized that it had been discovered. Expecting it to rise from the sand and swim away, behavior we had seen many times before, we maintained our position. However, this octopus swam straight toward us, at speed.

We tried to back down, but the octopus lunged at our camera and ripped the diving regulator out of author M.E.G.'s mouth. With amusement and then growing concern, author J.H.-U. watched as M.E.G. tried to free the regulator (which supplies a diver's breathing air) from the octopus's arms. But the animal refused to let go.

J.H.-U. entered the fray, sharing air with M.E.G. and helping him wrestle with the octopus to retrieve the regulator, now entangled in multiple arms and held fast by suction. To our relief, after a few seconds, the octopus let go of the equipment. As it did, it released a cloud of black ink into the water to distract us and ensure its quick escape. Shaken, we watched as it disappeared, unharmed, into the sea. Our octopus encounter underscored the importance of creating MPAs to keep these magnificent creatures safe from illegal fishing and overfishing. Granting them the protection they deserve will also decrease the chances of interaction with humans—we would hate for a fellow diver to misinterpret their behavior and accuse them of being multi-armed and dangerous!

Call for submissions
Life in Science is an occasional feature highlighting some of the humorous or unusual day-to-day realities that face our readers. Can you top this? Submit your story to www.submit2science.org.

Jorge Hernández-Urcera* and Manuel E. Garci
Department of Ecology and Marine Resources, Institute of Marine Research (IIM-CSIC), 36208 Vigo, Spain.
*Corresponding author. Email: jurcera@iim.csic.es

10.1126/science.add2639

Q&A with Mani L. Bhaumik

Bhaumik Breakthrough of the Year Award to recognize individual researchers

By **Andrea Korte**

Since 1996, the news and editorial staff of *Science* have looked back at the past year of scientific discovery to select the Breakthrough of the Year. In 2021, for instance, *Science* recognized the solution to a challenge that had puzzled scientists for 50 years: How to predict the shape that a string of amino acids will fold into as it becomes a working protein. In two papers published simultaneously in *Nature* and *Science* in 2021, researchers showed how this could be done using artificial intelligence-driven software. Other recent Breakthroughs include 2019's first-ever image of a black hole and 2020's effective vaccines against COVID-19.

Now, an international cash prize component—the largest bestowed by any scientific society—will recognize the individual scientists whose research and activities are foundational to the Breakthrough of the Year, AAAS announced in July. After the news and editorial staff of *Science* select and publish the *Science* Breakthrough of the Year, a committee convened by the *Science* journals' editor-in-chief will select up to three winners to receive the Mani L. Bhaumik Breakthrough of the Year Award. Starting in 2023, the convening of the selection committee will take place in January each year, and the announcement of the individual winner or winners will be made before the end of March.

Mani L. Bhaumik, a physicist whose work on laser technology paved the way for Lasik eye surgery, shares with AAAS and the *Science* family of journals a commitment to recognizing significant developments in scientific research. Bhaumik, whose gift is the largest in AAAS history, spoke with AAAS News & Notes about his inspirations and his current projects.

Q: Why was it important to you to enhance the Breakthrough of the Year with a new prize component that recognizes individual scientists?

A: The Breakthrough of the Year is done by people—it just doesn't fall from the sky. People are really at the center of the Breakthrough.

I am still a practicing scientist at 91 years old, and I know that beyond sheer curiosity about our world, peer recognition is one of the things that drives us. To be able to recognize hard-working scientists with this award—and perhaps prompt even more good work—is exciting.

There are so many researchers focused on every aspect of every subject being considered for the Breakthrough of the Year. Choosing those scientists most integral to the winning research will be a big and challenging decision for the selection committee to make.

Q: Tell me about your connections with AAAS and *Science*.

A: I've been reading *Science* magazine since I started my career, when I came to California in 1959 and was working on laser development. When you are in such an active field, you have to run like crazy and know the latest updates. But *Science* also covers such diverse subjects. You get a glance of what's important across many fields.

In 2019, I also endowed a prize at AAAS to support science communication and recognize excellent science communicators, the AAAS Mani L. Bhaumik Award for Public Engagement with Science. Many people think science is too esoteric. And even if they're interested in the results, they might not understand the processes unless it is explained well. But it's important for people to have that information in a way they can understand. Scientific knowledge is not just for scientists. It should be useful to everybody.

Some of my recent projects are focused on reaching audiences and sharing science with them in a meaningful way that makes sense to them. Recently I created a cartoon series for children called *Cosmic Quantum Ray*. I created a villain who breaks natural laws—he steals gravity. And the good guy, Quantum Ray, comes in with his crew and restores the gravity. It shows children how natural phenomena like gravity are all around us. I'm now working on a movie for adults about my main research interest: quantum field theory and how it shows that we're all connected. There's magic in moviemaking to get people to pay attention.

Q: Among your projects is the Bhaumik Educational Foundation, which allows students in rural West Bengal, India, to attend college for free. Why is it important for you to give back?

A: I came from extreme poverty—I didn't have any shoes until I was 16. The nearest high school was about 4 miles away from my village. So, I walked there every day and was lucky to have good teachers and fall in love with mathematics and science. To go to college, I went to Kolkata and did private tutoring to support myself and my studies. When I went back to India for a visit in 2000, I read a newspaper article that among the top high school graduates, at least half of them have no resources to fulfill their dreams. I thought, this is the best way I can pay off my dues from the people who helped me. I started the foundation to give back to my country of origin, and it's put about 300 students so far through university studying the science, engineering, math, and medicine that India needs.

Q: What were the breakthroughs and who were the researchers who inspired you during your own research career?

A: Satyendra Nath Bose was my mentor and teacher. He got me interested in theoretical physics. And Paul Dirac came to visit him—Dirac is known as the father of the quantum field theory. He started talking about that, and to me it sounded unbelievable, that whether matter or force, they all come from similar types of fields. Meeting Dirac was a dramatic experience. His talk was something I never had thought about: All electrons are exactly the same throughout the universe, and there is a common origin.

Q: What wisdom or advice would you give to the future winners of the Breakthrough of the Year?

A: My humble advice to the potential winners of the award would be for them to realize that science is a very hard taskmaster and demands dedication with unremitting strenuous work for success. Fortunately, the ecstasy of revealing the secrets of nature makes the arduous work feel like fun.

This interview has been condensed and edited for clarity.

PHOTO: MANI L. BHAUMIK

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

PHOTO: DANIEL BOSMA/GETTY IMAGES

COASTAL ECOLOGY

Will marshes rise up or sink?

Marsh ecosystems are vulnerable to rising sea level, in addition to land-use change and other human activities. Studies have shown that some marshes are gaining elevation, making them remarkably resilient to rising seas; however, results vary across locations and between contemporary and Holocene records. Comparing data from 97 sites on four continents, Saintilan *et al*. found that the relationship between sediment accretion and marsh subsidence explains the variable responses to sea-level rise. Marshes accrete more sediment, keeping up with sea-level rise up to a point, but sediment subsidence increases nonlinearly with accretion such that at higher rates of sea-level rise, marshes begin to sink. Marshes are unlikely to keep up with rising seas under current climate change projections. —BEL

Science, abo7872, this issue p. 523

Tidal marshes, such as those found in the Wadden Sea in the Netherlands, will likely not accumulate sediment fast enough to keep up with rising waters.

SOLID-STATE IONICS

Speed limit for solid-state nanoionics

The speed of biological information processing in neurons and synapses is limited by the aqueous medium through which weak action potentials of about 100 millivolts propagate over milliseconds. Above 1.23 volts, liquid water decomposes. Artificial solid-state neurons are not limited by such time and voltage constraints and can also be fabricated at the nanoscale, 1000 times smaller than their biological counterparts. Using complementary metal-oxide semiconductor–compatible materials, Onen *et al*. prototyped nanoscale protonic programmable resistors that can withstand high electric fields of around 10 megavolts per centimeter and which have energy-efficient modulation characteristics at room temperature. The proposed devices are 10,000 times faster than biological synapses and offer a promising direction for implementing various applications that can benefit from fast ionic motion. —YS

Science, abp8064, this issue p. 539

CANCER

Targeting drug-resistant brain cancer

A glioma is a tumor of the glial cells found in the brain and spinal cord. Glioblastoma multiforme (GBM) is the most common type of glioma, and is a highly aggressive brain tumor in dire need of new treatment strategies. GBM is frequently treated with the chemotherapy drug temozolomide (TMZ), but resistance develops in more than half of patients. Using a medical chemistry approach, Lin *et al.* designed TMZ analogs to overcome drug resistance in GBM (see the Perspective by Reddel and Aref). These agents generate a primary DNA lesion that can be repaired by healthy cells with intact DNA repair mechanisms. However, cancer cells that lack DNA-repair machinery are not able to repair the damage and

slowly evolve to harbor more toxic secondary lesions that cause selective tumor cell killing. —PNK

Science, abn7570, this issue p. 502; see also add4839, p. 467

CILIA AND FLAGELLA

Close-up view of transport train assembly

Cilia and flagella extend from the cell surface of various eukaryotic cells and perform diverse motility and signaling functions. The base of the cilium controls the entry of large intraflagellar transport trains that carry important cargo proteins throughout this specialized organelle. Defects in the ciliary base result in altered ciliary composition and human diseases. van den Hoek *et al*. used a combination of cryo–electron tomography and expansion microscopy techniques to study the molecular architecture of the ciliary base in the green alga *Chlamydomonas reinhardtii*. Their findings elucidate how intraflagellar transport trains assemble before they enter cilia and demonstrate the possibility of visualizing dynamic events with molecular resolution inside native cells. —SMH

Science, abm6704, this issue p. 543

NEURODEVELOPMENT

Life's bumpy road

Throughout life, cells of the body acquire somatic mutations. Bae *et al*. took a close look at the mutational load on the human brain. A survey of genomic alterations from normal brain samples and samples from brains affected with Tourette syndrome, schizophrenia, or autism spectrum disorder showed that although most human brains do acquire these somatic mutations, a subset seem to be hypermutable. In the samples of brains from people with autism spectrum disorder, some of these mutations created greater opportunities for transcription factor binding. —PJH

Science, abm6222, this issue p. 511

MATERIALS SCIENCE

Correlative straining mapping

The ability to simultaneously track elastic, plastic, and total strains with high resolution during the deformation of engineering materials is vital to understanding the micromechanisms of materials mechanics. Studying single-crystal tungsten under mechanical load, Edwards *et al.* demonstrated the feasibility of measuring pure elastic and plastic components of deformation with submicrometer resolution by combining high-resolution electron backscatter diffraction and nanoscale digital image correlation. These measurements shed new light on toughness-inducing plasticity at a crack tip in tungsten. This approach can be used to study the elastic and plastic processes relevant to microstructure in various types of materials. —ML

Sci. Adv. 10.1126/sciadv.abo5735 (2022).

SKIN INFLAMMATION

Squalene incites acne

Lipids from the skin's sebaceous glands contribute to the pathogenesis of acne, but the molecular circuits allowing lipids and skin bacteria to collaborate and incite lesion formation are unclear. Do *et al.* used single-cell analysis technologies, including spatial transcriptomics, to show that many macrophages in acne lesions belong to the TREM2 (triggering receptor expressed on myeloid cells 2) subset and are found adjacent to hair follicles and sebaceous glands. The addition of squalene to monocyte-derived macrophages in culture promoted their differentiation into TREM2 macrophages that exhibited reduced killing of *Cutibacterium acnes* bacteria. These findings identify disease-associated proinflammatory pathways that can be therapeutically targeted as part of an effort to better control the appearance of new lesions in patients with chronic acne. —IRW

Sci. Immunol. **7**, eabo2787 (2022).

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

ANIMAL BEHAVIOR

Individuality in avian predators

In an ergodic system, the average behavior of an individual over time resembles the average behavior of many individuals at a single time. The ergodicity of a system, or the lack thereof, is important for understanding any large-scale phenomenon such as the stock market or the migratory patterns of species. Vilk *et al.* tracked a total of 70 birds, all avian predators, including barn owls, black-winged kites, and common kestrels, and found that they exhibited ergodic behaviors when cruising over long distances but nonergodic behaviors when foraging over short distances. The authors believe that the nonergodic hunting preferences of the individual birds are an evolutionary adaptation to how their prey vary their survival strategies. —YY *Phys. Rev. X* **12**, 031005 (2022).

The hunting preferences of the barn owl (shown here) and other birds are an evolutionary adaptation to how their prey vary their survival strategies.

PHOTO: GUY EDWARDES/MINDEN PICTURES

NANOMATERIALS

Reusing solvent in nanoparticle synthesis

Colloidal synthesis of metal nanoparticles can create tighter size distributions and potentially greater control of reactivity in heterogeneous catalysis than conventional methods such as wet impregnation. However, colloidal synthesis is a costly alternative because solvents and surfactants are normally used only once. Wrasman *et al*. added an antisolvent mixture of 2-propanol and ethanol to precipitate the nanoparticles after colloidal synthesis and then recovered the hydrocarbon solvent and unused surfactant through distillation. Synthesis of metal nanoparticles such palladium and silver, and alloys such as platinum-tin, could be repeated up to 10 times with the recovered materials while maintaining the initial size distributions. These nanoparticles delivered highly reproducible results in reactions such as propane dehydrogenation and may offer a cost-competitive alternative as catalysts. —PDS

J. Am. Chem. Soc. **144**, 11646 (2022).

PLANT SCIENCE

Legacies of drought

Drought affects aboveground plants, as well as their belowground associated microbial and fungal communities, to influence plant growth. Lozano *et al.* show that soil microbial communities carry the effects of previous drought even into water-replete situations. A variety of grassland species, including graminoids, forbs, and legumes, were grown in soil that had previously hosted plants in drought or well-watered conditions. Root morphology showed changes characteristic of water scarcity when plants were

grown in soil that had previously experienced drought even though the plants themselves had not. Soil that has experienced drought may be deficient in mycorrhizal fungal mutualists. As a result, plants growing in drought soil may grow thicker roots to encourage the establishment of mutualistic associations from depleted soil. —PJH

New Phytol. 10.1111/nph.18327 (2022).

ENVIRONMENT

Managing lake water quality

Large bodies of freshwater are under increasing threat from climate change and human activities. The clear waters of Lake Geneva risk eutrophication and declining deepwater oxygen because of organic matter and phosphorus inputs that promote algal growth. Deyle *et al.* applied machine learning to time-series data for Lake Geneva to gain a more precise understanding of the variables affecting water quality. There have been efforts to reduce phosphorus inputs, but success has been limited, in part because increasing ambient temperature also promotes algal growth. Interdependencies among biogeochemical variables were included in a machine-learning approach coupled with conventional models for predicting deep lake–dissolved oxygen levels. The combined approach offers more nuanced guidance for lake water quality management. —CA

Proc. Natl. Acad. Sci. U.S.A. **119**, e21024661 (2022).

STEM WORKFORCE

Women choosing greener pastures

Women faculty remain underrepresented in chemistry because of a lack of applicants. Howe *et al*. surveyed 130 women and found that themes of overwork, unrealistic expectations, and an emphasis on research over teaching and mentoring were the underlying reasons that women looked elsewhere for employment. Further analysis revealed that women felt unsupported during graduate school and extrapolated this lack of support to faculty positions. This suggests that graduate-level experiences inform interests and career goals, and that systemic changes for increasing the diversity of faculty need to be made at the graduate level. Collectively, the results show that diversifying faculty positions is much more than simply opening a position. Instead, it requires policy changes that explicitly value women and their ideals. —MMc

JACS Au. **2**, 1443 (2022).

PANCREATIC CANCER

Exercise boosts tumor immunity

Exercise has many positive effects on the body, including improved blood circulation, heart strengthening, and weight control. Kurz *et al.* report that aerobic exercise is also associated with reduced tumor growth in models of pancreatic ductal adenocarcinoma. They showed that exercise enhanced the survival of a type of immune cell called $CD8^+$ T lymphocytes, which are known for their ability to kill tumors. Greater numbers of $CD8^+$ T cells capable of infiltrating tumors were found to accumulate as a result of an exercise regimen. The antitumor effects were mediated through the interleukin-15 (IL-15) axis, and pharmacological targeting of the IL-15/IL-15Rα pathway slowed tumor growth, particularly when combined with immunotherapy and chemotherapy treatments. —PNK

Cancer Cell **40**, 720 (2022).

STEM PUBLISHING

Gender inequality

Do inequities in women's editorships associate with an underrepresentation of women authors? Last *et al.* examined the association between women's editorship and women's authorship across randomly selected journals in infectious diseases spanning each impact factor quartile. Only 33% of infectious disease journals' editors-in-chief were found to be women. While controlling for journal impact factor, analysis of over 10,000 citable infectious diseases articles found that although nearly 50% of first authors were women, only 35% of last authors were women. Moreover, the presence of more women editors was associated with more women first authors, as well as more women last authors. —EEU

Lancet Infect. Dis. 10.1016/S1473-3099(22)00367-X (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

INNATE IMMUNITY

A common core of immune responses

Dealing with pathogens is a constant struggle for many life-forms, and innate and adaptive immune systems are needed to support survival. Essuman *et al*. reviewed the latest insights into the Toll/interleukin-1 resistance/receptor (TIR) domain proteins, which support immune responses across the tree of life, from archaea to bacteria to plants to humans. TIR domains function in innate immune signaling pathways, as well as in axon degeneration in animals. Some, but not all, TIR domains have enzymatic activity. TIR nicotinamide adenine dinucleotide hydrolase activity starves phages as they infect prokaryotes and promotes hypersensitive cell death in a plant's response to pathogens. Plant TIR domains can also synthesize small, nucleotide-based second messengers that initiate an immune response. —PJH

Science, abo0001, this issue p. 486

SOLAR CELLS

A radical doping approach

In perovskite solar cells, high power conversion efficiencies (PCEs) are usually obtained with an organic hole transporter called spiro-OMeTAD. This material must be doped to have sufficient conductivity and optimal work function, but the conventional process with lithium organic salts requires a long oxidation step that also affects device stability. Zhang *et al*. added spiro-OMeTAD biradical precursors that convert into stable organic monoradicals. Combined with ionic salts, this doping strategy formed solar cells with high PCEs (>25%) and improved stability. This approach also allows conductivity and work function to be tuned separately and could be applied in other optoelectronic devices. —PDS

Science, abo2757, this issue p. 495

BIOPHYSICS

A new force in chromatin research

Fundamental questions about the physical nature of chromosomes remain unanswered, largely due to the absence of direct mechanical measurements inside the nuclei of living cells. Keizer *et al*. developed a technique to measure how a genomic locus, in its native nuclear context, responds to a point force of physiological magnitude (see the Perspective by So and Tanner). Interphase chromatin was found to be liquid like, with moderate hindrance and topological effects, contrasting with the common view of a crowded and entangled nuclear environment. Ultimately, these measurements allow a deeper understanding of how genomic elements can move under active biological forces and provide a basis for developing new physical models of chromosomes. —DJ

Science, abi9810, this issue p. 489; see also add5444, p. 472

BIOIMAGING

Continuous long-term ultrasound imaging

Ultrasound is widely used for the noninvasive imaging of tissues and organs, but this method requires close contact between the transducer and the target area. This can make it difficult to acquire images over a long period of time, especially if the patient needs to be mobile. Wang *et al*. describe a wearable ultrasound imaging device (see the Perspective by Tan and Lu). A rigid piezoelectric probe array is bonded to the skin with an acoustically transparent hydrogel elastomer. In vivo testing showed that the device could be comfortably worn for 48 hours, and hooking the array up to a commercially available ultrasound platform allowed for continuous ultrasound images of the carotid artery, lung, and abdomen. —MSL

Science, abo2542, this issue p. 517; see also adc8732, p. 466

SPECIES INTERACTIONS

"Pollination" of red algae

Pollination by animals is a widespread and important symbiotic relationship that increases plants' genetic diversity and fertilization success. Lavaut *et al*. discovered that this relationship is not unique to plants; it is also found in the dioecious red alga *Gracilaria gracilis* (see the Perspective by Ollerton and Ren). The isopod *Idotea balthica* carries the alga's spermatia on its body, and experiments showed that isopods moving from a male to a female alga increased the alga's fertilization. Transport of algal gametes by isopods could be an important mechanism for reproduction because the spermatia lack flagella and otherwise rely on water currents to transport them. —BEL

Science, abo6661, this issue, p. 528; see also add3198, p. 471

MOLECULAR BIOLOGY

Recording events in cells

CRISPR-mediated DNA editing has turned toward recording events in cells, allowing cell differentiation lineages to be determined in vivo and potentially enabling gene expression and cell signaling to be recorded. These CRISPR information recorders rely on mutations being introduced to synthetic "DNA tapes" that are expressed in cells and can be sequenced after tissue development and growth to reconstruct changes over time. In a Perspective, Masuyama *et al.* discuss the potential of this approach to better understand the development and pathogenic progression of tissues such as tumors. Technological challenges currently limit the application of these systems, but improvements are being developed rapidly. —GKA

Science, abo3471, this issue p. 469

CARBON NANOTUBES

Ordering surface modifications

Although several methods exist for functionalizing the surface of carbon nanotubes, the sites modified tend to be random. However, ordered functionalization of single-walled carbon nanotubes (SWCNTs) could lead to quantum properties. Lin *et al*. screened short DNA sequences containing cytosine (C) and guanine (G) to control the placement of G for photo–cross-linking reactions (see the Perspective by Wang). Spectroscopy and cryo–electron microscopy showed that for the (8,3)-SWCNT enantiomer, the sequences $C_3GC_7GC_3$ regularly created an ordered helical structure with a 6.5-angstrom periodicity in which G bridges two carbon atoms. —PDS

Science, abo4628, this issue p. 535 see also abq2580, p. 473

PARKINSON'S DISEASE

Microglia modulates PD risk

Previous studies have identified variants in the *LRRK2* gene that are associated with increased risk of developing Parkinson's disease (PD). Langston *et al.* studied the mechanisms linking *LRRK2* variants to PD and showed that the disease risk is determined by microglial cells. Studying PD-associated *LRRK2* variants in patient-derived tissue and induced pluripotent stem cell–derived microglia, the authors identified a quantitative trait in microglia, suggesting that although this cell type expresses *LRRK2* in lower amounts compared with other cell types, it might be critical for determining PD risk. The lack of correlation between

cell-type-specific baseline expression and effect on disease pathophysiology should be considered when evaluating the mechanisms mediating this and other disease-associated variants in noncoding gene regions. —MM

Sci. Transl. Med. **14**, eabp8869 (2022).

PHYSIOLOGY

A painful breakdown of cartilage

The proinflammatory cytokine interleukin-6 (IL-6) is associated with inflammatory pain and structural changes in joints that characterize osteoarthritis. Using a mouse model of injury-induced osteoarthritis, Liao *et al*. demonstrated that IL-6 promoted both disease symptoms in a sex-specific manner (see the Focus by Pacifici). In male mice, but not in female mice, IL-6 promoted articular cartilage degeneration and pain signaling through distinct signaling pathways. These findings have implications for potential sex-specific osteoarthritis therapies. —AMV

Sci. Signal. **15**, eabn7082, eadd3702 (2022).

PLANT SCIENCE

Nucleotide-based signals protect

Key receptors in flowering plants respond to pathogen effector molecules and trigger the plant's immune response when needed. Huang *et al*. and Jia *et al*. now tell the story of a branching signaling pathway through which receptors with an N-terminal Toll/interleukin-1 receptor (TIR) domain enzymatically produce small, nucleotide-based second messengers and initiate immune responses in the small mustard plant *Arabidopsis*. The pathways diverge to activate different immune responses, with one pathway controlling the cell death response and the other managing transcriptional change. —PJH

Science, abq3297, abq8180, this issue p. 487, 488

SOLAR CELLS

Managing excess lead iodide

In hybrid perovskite solar cells, the formation of lead iodide (PbI_2) can provide some passivation effects but can lead to device instability and hysteresis in current-density changes with voltage. Zhao *et al*. show that doping with rubidium chloride (RbCl) can create a passive inactive $(PbI_2)_2RbCl$ phase that stabilizes the perovskite phase and lowers its bandgap. Devices exhibited 25.6% certified power efficiency and maintained 80% of that efficiency after 500 hours of operation at 85°C. —PDS

Science, abp8873, this issue p. 531

REVIEW SUMMARY

INNATE IMMUNITY

Shared TIR enzymatic functions regulate cell death and immunity across the tree of life

Kow Essuman, Jeffrey Milbrandt, Jeffery L. Dangl, Marc T. Nishimura*

BACKGROUND: Diverse organisms, from archaea and bacteria to plants and humans, use receptor systems to recognize both pathogens and dangerous self-derived or environmentally derived stimuli. These intricate, well-coordinated immune systems, composed of innate and adaptive components, ensure host survival. In the late 20th century, researchers identified the Toll/interleukin-1/resistance gene (TIR) domain as an evolutionarily conserved component of animal and plant innate immune systems. Today, TIR-domain proteins are known to be broadly distributed across the tree of life. The TIR domain was first recognized as an adaptor for the assembly of macromolecular signaling complexes in mammalian innate immune pathways. Work on axon degeneration in animals—as well as on plant, archaeal, and bacterial immune systems—has uncovered additional enzymatic activities for TIR domains.

ADVANCES: Mammalian axons initiate a self-destruct program upon injury and during disease that is mediated by the sterile alpha and TIR motif containing 1 (SARM1) protein. The SARM1 TIR domain enzymatically consumes the essential metabolic cofactor nicotinamide adenine dinucleotide (NAD^+) to promote axonal death. Identification of the SARM1 NAD^+-consuming enzyme (NADase) revealed that TIR domains can function as enzymes. Given the evolutionary conservation of TIR domains, studies investigated whether the SARM1 TIR NADase was also conserved. Indeed, bacteria, archaea, and plant TIR domains possess NADase activity. In prokaryotes, TIR NADase activity is found in an ancient antiphage immune system. In plants, identification of TIR NADase activity and linkage of TIR enzymatic products to downstream signaling components addressed the question of how nucleotide-binding, leucine-rich repeat (NLR) receptors trigger hypersensitive cell death during an immune response. Studies in plants show that their TIR domains can cleave nucleic acids and possess 2′,3′ cyclic adenosine monophosphate (2′,3′-cAMP) and 2′,3′ cyclic guanosine monophosphate (2′,3′-cGMP) synthetase activity that aids cell death programs in plant innate immunity. Thus, TIR domains constitute an ancient family of enzymes that are activated in immune and cell death pathways.

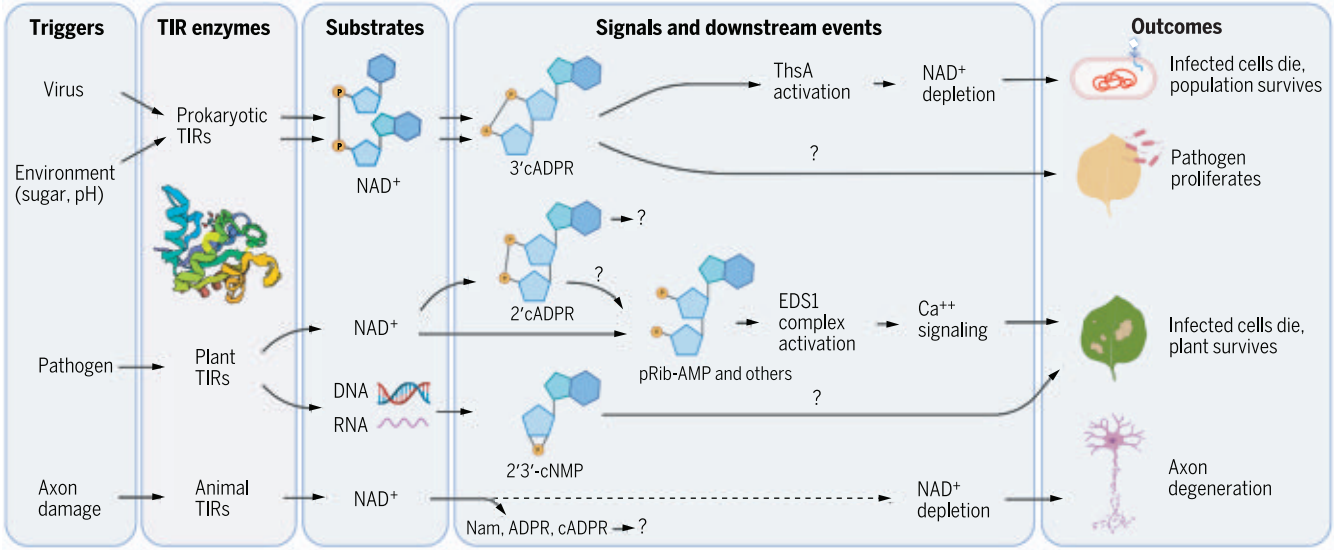
OUTLOOK: The discovery of TIR-domain enzyme activities carries implications for innate immunity and neurodegeneration. The identification of the SARM1 NADase defined a drug target for a wide number of neurodegenerative diseases that is being exploited in both preclinical and clinical studies. Hyperactive mutations in the SARM1 NADase have been discovered in amyotrophic lateral sclerosis (ALS) patients. Future work will seek to clarify the contribution of the SARM1 axon degeneration pathway to ALS pathogenesis. NAD^+ biology influences cellular processes from metabolism to DNA repair to aging. How TIR enzymes influence the NAD^+ metabolome and its associated pathways in bacteria, archaea, plants, and animals will be an exciting area for upcoming investigation. The discovery of the diversity of TIR enzymatic products is revealing signaling pathways across kingdoms. Discovery of TIR enzymatic function in plants and animals may yet inspire studies of enzymatic functions for Toll-like receptors in animals. We anticipate that cross-kingdom studies of TIR-domain function will guide interventions that will span the tree of life, from treating human neurodegenerative disorders and bacterial infections to preventing plant diseases. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: marc.nishimura@colostate.edu

Cite this article as K. Essuman *et al.*, *Science* **377**, eabo0001 (2022). DOI: 10.1126/science.abo0001

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abo0001



Conserved TIR-domain enzymatic activity. TIR-domain proteins from prokaryotes and eukaryotes cleave NAD^+ into nicotinamide (Nam), ADP-ribose (ADPR), cyclic ADP-ribose (cADPR), isomers of cyclic ADP-ribose (2′ or 3′cADPR), and related molecules [e.g., phosphoribosyl adenosine monophosphate (pRib-AMP)]. Plant TIR domains also possess a nuclease activity, can degrade DNA and RNA, and can function as a 2′,3′-cAMP or 2′,3′-cGMP synthetase. TIR enzymatic activity drives cell death and immune pathways across kingdoms. TIR activity can kill cells directly through NAD^+ depletion or indirectly using enzymatic products as signal molecules. The representative TIR domain structure shown here is Protein Data Bank ID 6O0Q. EDS1, enhanced disease susceptibility 1; ThsA, Thoeris A.

REVIEW

INNATE IMMUNITY

Shared TIR enzymatic functions regulate cell death and immunity across the tree of life

Kow Essuman[1], Jeffrey Milbrandt[2,3,4], Jeffery L. Dangl[5,6], Marc T. Nishimura[7]*

In the 20th century, researchers studying animal and plant signaling pathways discovered a protein domain that is shared across diverse innate immune systems: the Toll/interleukin-1/resistance gene (TIR) domain. The TIR domain is found in several protein architectures and was defined as an adaptor that mediates protein-protein interactions in animal innate immunity and developmental signaling pathways. However, studies of nerve degeneration in animals—and subsequent breakthroughs in plant, bacterial, and archaeal systems—revealed that TIR domains possess enzymatic activities. We provide a synthesis of TIR functions and the role of various related TIR enzymatic products in evolutionarily diverse immune systems. These studies may ultimately guide interventions that would span the tree of life, from treating human neurodegenerative disorders and bacterial infections to preventing plant diseases.

Innate immune systems of living organisms comprise a set of molecular, cellular, and physiological responses that are rapidly mobilized in response to pathogens and harmful stimuli derived from damaged host cells (*1*, *2*). Studies of animal and plant innate immune responses identified the evolutionarily conserved Toll/interleukin-1/resistance gene (TIR) domain as a common element in proteins involved in innate immunity (Fig. 1) (*3*). In animal innate immune systems, TIR domains provide the signaling motif of Toll-like receptors (TLRs) and interleukin-1 receptors (IL-1Rs) (*4*). These TIR domains function as scaffolds to assemble signaling complexes and transduce defense responses that include activation of nuclear factor κB (NF-κB) signaling and interferon production (*4*). Activation of mammalian TIR protein complexes also stimulates aerobic glycolysis in functional immune cells (*5*, *6*). In plants, TIR domains are typically found at the amino terminus of nucleotide-binding leucine-rich repeat receptor (NLR) proteins (Fig. 1) (*7*, *8*). Canonical plant NLRs possess a nucleotide binding site (NBS), leucine-rich repeat (LRR) domain, and a TIR domain or coiled-coil (CC) domain at their N terminus, giving rise to TIR-NLRs (TNLs) and CC-NLRs (CNLs), respectively. NLRs are the key intracellular receptors in the plant innate immune system (*7*, *8*) and are structurally similar to animal NLR proteins that signal by oligomerizing into inflammasomes (*7*). In response to pathogenic and endogenous danger signals, plant NLRs trigger a rapid immune response that initiates a transcriptional reprogramming that halts pathogen growth and often culminates in localized cell death (*9*). TIR domains are also encoded (more rarely) in some fungal and protozoal genomes where their function remains to be elucidated (*10*, *11*). More recently, TIR domain proteins were discovered to be widely found in bacteria and archaea (Fig. 1) (*12*–*17*). Initial characterization of these bacterial proteins indicated that they function as virulence factors to interrupt TLR signaling in the mammalian host, but recent work also defines intracellular signaling roles in bacterial antiphage defense (*12*, *13*, *16*, *18*). Thus, TIR domain immune function is found across kingdoms in diverse systems.

The original molecular function of TIR domains as signaling scaffolds is well documented for innate immune responses in animals (*4*, *19*–*21*). Recent work, however, demonstrated that TIR proteins are also an ancient enzyme family across all domains of life (*15*, *18*, *22*–*26*). TIR domains can act as nicotinamide adenine dinucleotide (NAD^+) hydrolases (*15*, *22*–*26*). Additionally, at least some plant TIR domains are bifunctional enzymes that are also able to hydrolyze both NAD^+ and DNA and RNA, thus also acting as 2′,3′ cyclic adenosine monophosphate (2′,3′-cAMP) and 2′,3′ cyclic guanosine monophosphate (2′,3′-cGMP) synthetases (*23*, *24*, *27*). In this review, we discuss the biology of TIR domains with an emphasis on enzymatic functions in immune and cell death pathways (summarized in Fig. 1B).

[1]Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [2]Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA. [3]Needleman Center for Neurometabolism and Axonal Therapeutics, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA. [4]McDonnell Genome Institute, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA. [5]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [6]Howard Hughes Medical Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [7]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA.
*Corresponding author. Email: marc.nishimura@colostate.edu

Discovery of the TIR domain

The identification of the TIR domain began with the characterization of three key genes that encode the mammalian IL-1R, *Drosophila* Toll, and the tobacco N protein, which conditions resistance to tobacco mosaic virus (*19*). In the late 1980s, IL-1R was cloned (*28*). A few years later, its cytosolic domain was found to be homologous to the cytosolic domain of *Drosophila* Toll (*29*, *30*). Toll, at that time, was implicated in the dorsoventral patterning developmental process of *Drosophila* that involved signaling by the NF-κB factor Dorsal (*31*, *32*). IL-1 was also shown to activate NF-κB (*29*, *33*). In 1994, researchers studying the plant immune system discovered that the gene *N* encodes an amino-terminal domain with substantial homology to the carboxyl-terminal domain of Toll and IL-1R (*34*). Because of its N-terminal TIR domain, N is a TNL receptor. Its activation by recognition of a protein from the tobacco mosaic virus leads to hypersensitive programmed cell death and restriction of viral replication (*34*). Functions for Toll were later extended beyond development to immune activation during bacterial and fungal infections (*35*–*37*). These discoveries led to the identification of several mammalian TLR proteins harboring TIR domains and a large superfamily of intracellular plant TNL receptors that are widely distributed across dicotyledonous plant genomes (*4*, *7*). Single-domain plant TIR proteins and "TIR-X" proteins (where X is a domain of unknown function) are also common and functional (Fig. 1) (*38*–*40*).

TIR domains in animal immune signaling

The realization that TIR domains function as scaffolds emerged from extensive work in animal innate immune signaling (*19*). In animals, TIR domains are found in Toll and TLRs, their cytosolic adaptor proteins, and the IL-1R family (*4*, *19*) (Fig. 1). TLR receptors are pattern recognition receptors, which detect pathogen- or microbe-associated molecular patterns (PAMPs and MAMPs) and activate downstream signaling pathways (*4*). The detection of PAMPs or MAMPs occurs through the N-terminal domain of TLR dimers at the cell surface or within endosomal membranes (*4*). This triggers TLR C-terminal TIR domain dimerization, recruitment of cytosolic TIR adaptors (*4*, *21*, *41*), and TIR-mediated cooperative assembly of a supramolecular organizing complex (*19*–*21*, *42*, *43*). Complex formation leads to signaling through activation of downstream protein kinases and transcription factors, and then to host inflammatory

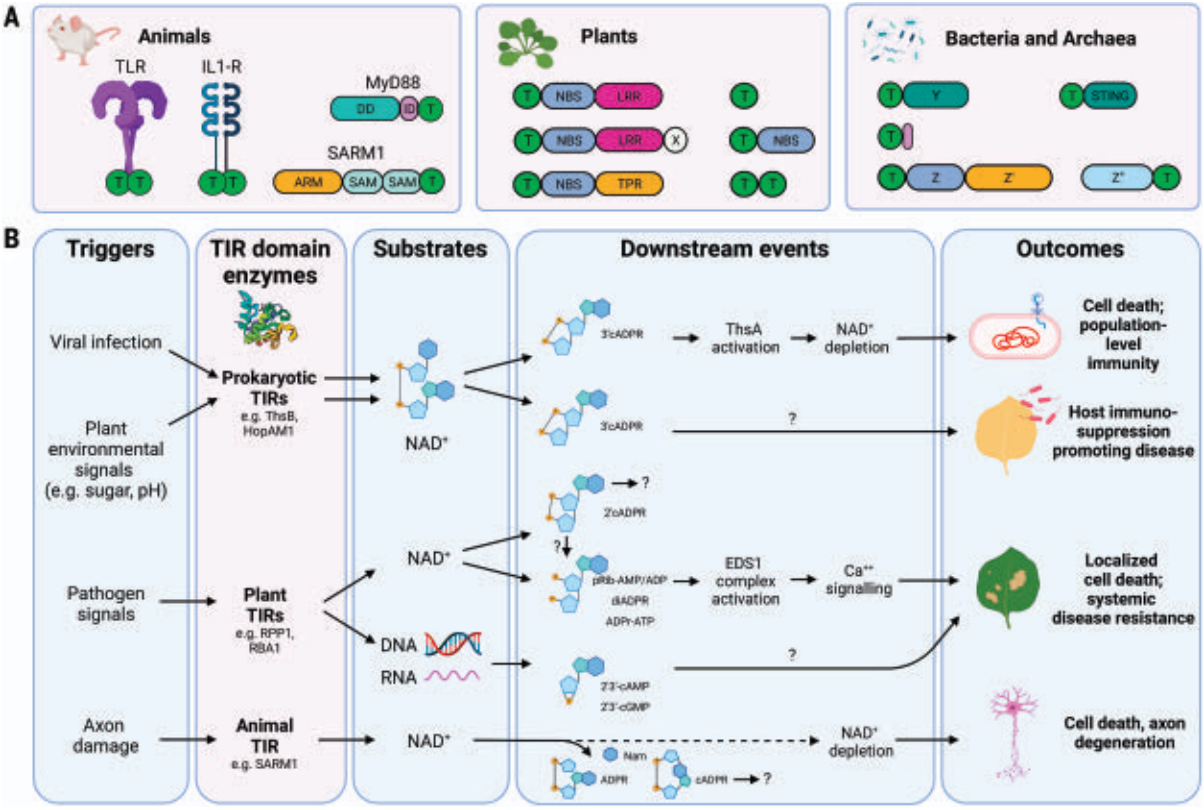


Fig. 1. Diverse TIR domain proteins across the tree of life share related enzymatic functions to regulate cell death and immunity. (**A**) TIR domains (labeled T, green circle) are present in diverse domain configurations from animal, plants, and prokaryotes. Many domain configurations are not shown. Plant genomes encode TIR domain proteins with diverse domain architectures. Bacteria and archaea also possess TIR domain proteins with diverse domain architectures, including TIR domains located at either the N or C termini. Other domains present in prokaryotic TIR proteins that are represented by Y-Z″ in the figure include CC domains, STING, WD 40 repeats, and the Mpr1p Pad1p N-terminal domain. DD, death domain; ID, intermediate domain; MyD88, myeloid differentiation primary response protein 88; TPR, tetratricopeptide repeat. (**B**) Across the tree of life, TIR domains function as enzymes to process nucleotide-containing molecules, either to generate signals or to deplete them from the cell. TIR products or NAD^+ depletion are often associated with beneficial cell death or immune activation.

and defense responses (*4*, *42*, *44*). TIR domains in the IL-1R family share similar intracellular signaling mechanisms as TLRs and influence myriad innate and adaptive immune responses (*4*, *45*–*47*).

Although animal TIR adaptor proteins generally promote TLR signaling, the discovery that the TIR "adaptor" protein, sterile alpha and TIR motif containing 1 (SARM1), inhibits TLR signaling indicated that it might have distinct functions (*48*). Indeed, the *Caenorhabditis elegans* SARM1 ortholog (called TIR-1) has essential roles in both immunity and development (*49*–*51*). TIR-1 is required for resistance to certain fungal and bacterial infections (*49*) and specification of asymmetric odorant receptor expression during neuronal development (*50*). Thus, *C. elegans* TIR-1 function is reminiscent of Toll function in dorsoventral patterning in development and innate immunity (*31*, *35*–*37*). Another distinguishing feature of SARM1 among TIR adaptors is its enzymatic role in driving axon degeneration (*22*, *52*, *53*) (Fig. 2), which laid the groundwork for the discovery of TIR domain enzymatic function across kingdoms.

TIR domains function in neuronal cell death

In the mid-19th century, neurophysiologist Augustus Waller described the degeneration of injured frog hypoglossal and glossopharyngeal nerves (*54*). Wallerian degeneration initially was thought to be a passive wasting of the distal segment of a damaged axon, but this view was challenged by the serendipitous discovery of a naturally occurring spontaneous mutant mouse, the *Wallerian degeneration slow* (*Wld*^{S}) mutant (*55*). In *Wld*^{S} mice, transected distal axons remain intact for several days after injury and continue to conduct action potentials and remain metabolically active (*55*, *56*). The axonal protection observed in the *Wld*^{S} mice is caused by a gain-of-function mutation that results in a fusion protein (Wld^S) composed of portions of the ubiquitination factor UBE4B and nicotinamide mononucleotide adenylyltransferase 1 (NMNAT1) (*56*, *57*). The enzymatic activity of NMNAT1, which converts nicotinamide mononucleotide (NMN) to NAD^+ (Fig. 3), is required for Wld^S to protect axons (*58*). Unraveling the mechanism of this hypermorphic mutation revealed an active, genetically encoded death program in axons after injury (*53*, *59*–*63*).

Genetic screens in *Drosophila* and in mouse dorsal root ganglia neurons identified SARM1 as a required component of a conserved axon death program (*61*, *62*). Subsequent experiments using SARM1 knockout mice showed robust axonal protection comparable to that observed in *Wld*^{S} mice (*56*, *61*, *62*). Transected axons from *Drosophila* SARM1 mutants can remain intact for at least 6 weeks, a duration that approximates the life span of the fly (*61*). Loss of SARM1 delays axon degeneration and improves functional outcomes in several neurodegenerative disease mouse models, including chemotherapeutic and diabetic peripheral neuropathy, traumatic brain injury, glaucoma, retinopathy, and amyotrophic lateral sclerosis (ALS) (*64*–*70*). Loss of SARM1 also blocks axon degeneration in human induced pluripotent stem cell–derived sensory neurons (*71*). In the absence of injury, SARM1 knockout mice display no overt

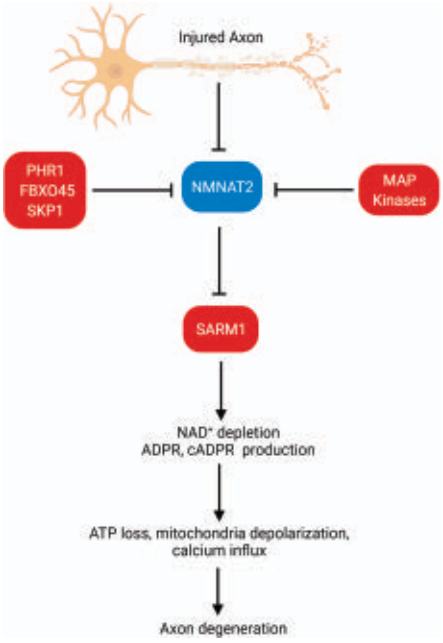


Fig. 2. Working model of the Wallerian axon degeneration pathway. Activation of the Wallerian axon degeneration pathway during injury involves the reduction of axon protective factors, primarily nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2). The level of NMNAT2 is regulated by multiple mitogen-activated protein (MAP) kinases, PAM-Highwire-Rpm-1 (PHR1), F Box protein 45 (FBXO45), and S phase kinase associated protein 1 (SKP1). Low levels of NMNAT2 lead to an increase in NMN and a decrease in NAD^+. Both NMN and NAD^+ bind to SARM1, where a high NMN:NAD^+ ratio is responsible for potent SARM1 NADase enzyme activation. Activated SARM1 NADase cleaves NAD^+ into Nam, ADPR, and cADPR. SARM1-mediated NAD^+ cleavage leads to energetic failure through axonal ATP loss and subsequent mitochondrial depolarization, calcium influx, and, ultimately, axon fragmentation.

phenotype, suggesting that SARM1 inhibition may be an attractive therapeutic strategy for neurodegenerative diseases (*72*–*74*).

SARM1 reveals an enzymatic TIR function

SARM1 is a multidomain protein with an N-terminal regulatory region that contains multiple Armadillo repeats, a tandem sterile alpha motif (SAM) domain that mediates SARM1 multimerization, and a C-terminal TIR domain (*52*, *62*) (Fig. 1). Forced dimerization of the SARM1 TIR domains leads to rapid NAD^+ depletion that is sufficient to trigger axon degeneration (*75*). Because TIR domains function as molecular scaffolds to promote the formation of signaling complexes, it was hypothesized that axon injury promoted dimerization of SARM1 TIR domains that resulted in the interaction and activation of an

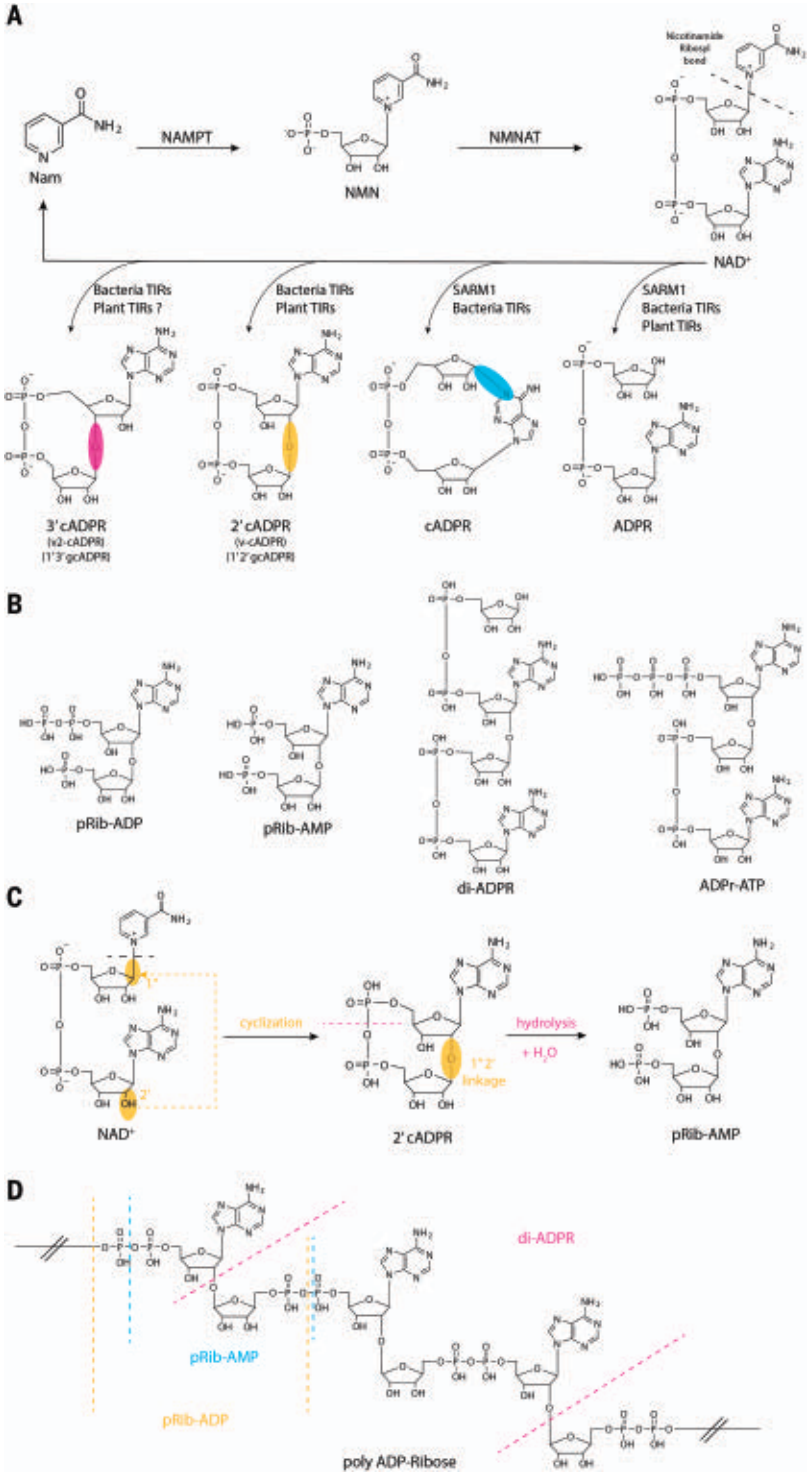


Fig. 3. Select pathways of NAD^+ synthesis and TIR-mediated NAD^+ degradation. (**A**) Nicotinamide phosphoribosyltransferase (NAMPT), the rate-limiting enzyme in NAD^+ biosynthesis, produces NMN from Nam and 5-phosphoribosyl-1-pyrophosphate (PRPP) (not shown). NMNAT synthesizes NAD^+ from NMN and ATP. NAD^+ can be cleaved by prokaryotic and eukaryotic TIR domains. The SARM1 TIR domain cleaves NAD^+ into Nam, ADPR, and cADPR. Bacterial TIRs can cleave NAD^+ into Nam, ADPR, cADPR, and v-cADPRs. Plant TIRs can also cleave NAD^+ into Nam, ADPR, and v-cADPR. Two linkages of v-cADPR have been described: 2′cADPR and 3′cADPR. (**B**) Plant TIRs generate small-molecule signals that activate the EDS1 complex. pRib-AMP and pRib-ADP promote the EDS1-PAD 4 interaction with ADR1, whereas ADPr-ATP and di-ADPR promote EDS1/SAG101 interaction with NRG1. (**C**) 2′cADPR and TIR-derived EDS1 signals share a 1″-2′ ribose-ribose linkage. (**D**) Hypothetical cleavage sites in poly(ADP-ribose) that could produce plant-derived TIR molecules.

NAD^+-consuming enzyme in the axon (*22*, *75*). However, biochemical experiments using purified proteins established that the SARM1 TIR domain itself possesses NAD^+ hydrolase activity (*22*). SARM1 hydrolyzes NAD^+ through a reaction that requires a glutamate residue [Glu^{642} (E642) in human SARM1] in the TIR catalytic pocket (*22*, *24*). This NAD^+ hydrolase activity and conserved glutamate is present in SARM1 homologs from multiple species, including humans, mice, flies, and worms (*22*, *24*, *26*).

SARM1 TIR cleaves NAD^+ at the nicotinamide ribosyl bond to produce nicotinamide (Nam), adenosine diphosphate ribose (ADPR), and cyclic ADPR (cADPR) (*22*, *24*, *26*) (Fig. 3). SARM1 also possesses base-exchange activity that can generate nicotinic acid adenine dinucleotide phosphate (NAADP) (*26*, *76*) as well as modifications of other compounds (*77*). SARM1 activation in axons leads to depletion of NAD^+ and adenosine triphosphate (ATP), followed by calcium influx, mitochondrial depolarization, and axon degeneration (*60*, *75*, *78*, *79*) (Fig. 2). The discovery of the SARM1 NAD^+-consuming enzyme (NADase) reemphasized the importance of NAD^+ in axon degeneration. Later studies showed that the Wld^S functional moiety, NMNAT1, confers axonal protection by inhibiting SARM1-mediated NAD^+ depletion (*80*).

Shared TIR enzymatic function across the tree of life

The discovery that the SARM1 TIR domain is an NAD^+ hydrolase suggested that TIR domains in proteins that are present in different kingdoms might also possess enzymatic activity. The SARM1 TIR domain is phylogenetically and structurally more related to bacterial, archaeal, and plant TIR domains than to TIR domains implicated in animal innate immunity (*14*, *24*). Indeed, purified TIR domain proteins of human TLR4 and other animal TIR adaptors failed to cleave NAD^+ in vitro (*22*, *24*, *81*). This result reinforced phylogenetic analyses and suggested that SARM1 may have functions distinct from other mammalian TIR domains. Biochemical assays using bacterial, archaeal, and plant TIR domains demonstrated that many of them can function as NADases (*15*, *17*, *18*, *23*–*26*). A conserved glutamate is present in plant, bacterial, and archaeal TIR domains and is required for their NAD^+ hydrolase activity (*15*, *16*, *23*, *24*). Other closely related analogs of NAD^+, such as nicotinamide adenine dinucleotide phosphate ($NADP^+$), can be cleaved by SARM1, plant, and bacterial TIR proteins (*15*, *23*, *24*, *76*). TLRs and their adaptors have not been shown to possess TIR enzymatic function thus far, but their ability to form complexes with other enzymes suggests that TIR-associated enzymatic activation may be a conserved mechanism across kingdoms (*6*, *19*, *42*). Moreover, TLRs can activate aerobic glycolysis and metabolic reprogramming (*5*, *6*, *82*). This, coupled with NAD^+ being an indispensable and ubiquitous metabolic factor (*83*), implicates TIRs as indirect mediators of metabolism. Future studies should clarify if mammalian TIRs that lack NAD^+ cleavage activity might still bind NAD^+ and thus also influence signaling pathways in this manner.

Bacterial and archaeal TIR domains are enzymes

TIR domain NADases are present and active in diverse prokaryotes from a range of habitats (*15*–*18*, *25*, *84*). Prokaryotic TIRs also cleave NAD^+ at the nicotinamide ribosyl bond (*15*, *25*) (Fig. 3) but can create multiple types of products. For example, TIR proteins from *Staphylococcus aureus* (TirS) and uropathogenic *Escherichia coli* (TcpC) generate Nam and ADPR (*15*). By contrast, TIR proteins BtpA (from *Brucella*), AbTir (from *Acinetobacter baumannii*), TcpO (from the archaea *Methanobrevibacter olleyae*), and other prokaryotic TIR domains generate a variant of cADPR (v-cADPR) (*15*, *18*, *84*) (Fig. 3). v-cADPR was also detected in gnotobiotic mice colonized with human gut bacterial strains (*17*). Plant TIRs also produce v-cADPR in vitro and in vivo (*23*). Finally, yet another variant of cADPR (v2-cADPR) is produced by the bacterial TIR virulence factor HopAM1 and other bacterial TIR proteins (*17*, *84*, *85*). The cADPR isomers produced by plant TIRs and several bacterial TIRs have the same molecular weight as cADPR but are distinct, as shown by retention times in chromatography experiments (*15*, *18*, *23*, *85*). The first chemical structures of plant and prokaryotic v-cADPR and v2-cADPR molecules indicate that they are circularized by means of ribose-ribose linkages (*84*, *86*) (Fig. 3). v-cADPR is circularized by means of a 1″-2′ glycosidic linkage between riboses (hereafter, 2′cADPR), whereas v2-cADPR is circularized by means of a 1″-3′ glycosidic linkage (hereafter, 3′cADPR) (*84*, *86*). It remains to be seen whether TIRs are capable of other linkages or whether there are isomeric differences between cADPR molecules that share the same linkage.

Bacterial TIR domain proteins were first classified as virulence factors and thought to act by interacting with mammalian host TLRs to disrupt innate immune pathways (*12*, *13*, *87*–*89*). BtpA is a bacterial virulence effector protein from *Brucella abortus* that reduces total intracellular NAD^+, an activity that requires the catalytic glutamate (*89*). Similar strategies are used by plant pathogens. The plant pathogen *Pseudomonas syringae* pv. (pathovar) tomato strain DC3000 encodes a TIR domain–containing virulence protein called HopAM1. HopAM1 virulence phenotypes are dependent on its conserved catalytic glutamate (*84*, *85*). The mechanism by which HopAM1's enzymatic activity suppresses plant immunity remains unknown. One hypothesis is that the HopAM1 TIR competes for a plant TIR substrate that is required to promote immunity. An alternate hypothesis is that HopAM1-produced 3′cADPR, or related enzymatic products, interfere with the host's immune system.

Bacterial TIR enzymes in antiphage defense responses

Global surveys of bacterial genomes revealed that TIR domain–encoding genes are enriched in genomic regions that contain phage defense genes, suggesting that they may function in suppressing phage infections (*16*). One system, Thoeris, comprises an operon that is composed of ThsA and ThsB. ThsA usually encodes an NAD^+ binding and cleavage domain at its N terminus (*18*, *90*). ThsB encodes a TIR-domain protein (*16*). The Thoeris system is found in a wide number bacterial and archaeal genomes and restricts viral replication upon activation (*16*). Mutations of the conserved catalytic glutamate in the TIR domain of ThsB, or of residues that affect NAD^+ binding in ThsA, abolish phage protection (*16*, *18*).

Upon bacteriophage infection, ThsB TIR domains produce a cADPR isomer (presumably 3′cADPR) (*18*, *84*), which then acts as a signaling molecule to activate ThsA (*18*). ThsA consequently consumes bacterial NAD^+, promoting bacterial growth inhibition and/or cell death (Fig. 4). This process of metabolic arrest and cell death, termed abortive infection, is thought to curb the spread of the infecting phage at a population level (*91*, *92*). Abortive infection is reminiscent of the axon self-destruct mechanism that is mediated by SARM1 activation in damaged axons and the hypersensitive cell death response in plants after successful pathogen recognition (*7*, *52*). The discovery that bacterial TIR domain–containing proteins mediate antiphage defense systems is an impressive example of the continuity of TIR immune function, extending from phage defense in bacteria to pathogen responses in plants and animals. In addition to ThsB, other phage defense proteins are TIR-domain NAD^+ hydrolases regulated by cyclic nucleotides (Fig. 4) (*25*, *93*). Similarly, detection of foreign DNA by a prokaryotic Argonaute protein via guide RNAs activates TIR NADase activity in the SPARTA (short prokaryotic Argonaute and TIR-APAZ) immune system (*94*).

TIR domains as enzymes in plant immune receptor function

Plants deploy an intricate innate immune system that prevents disease (*7*, *9*). Initial hypotheses imagined that similar to animal TIR proteins, plant immune receptors containing TIR domains could function through

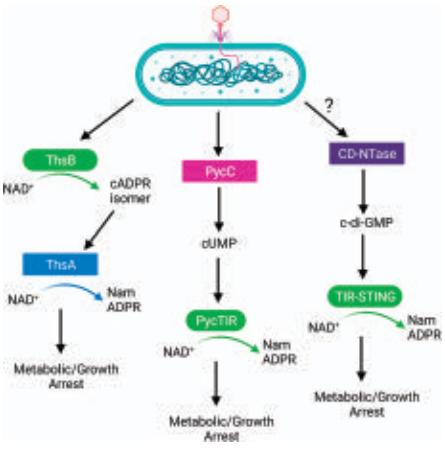


Fig. 4. Bacterial TIR NADases in antiphage, abortive infection defense programs. Shown are the Thoeris (left), Pycsar (middle), and CBASS (cyclic oligonucleotide–based antiphage signaling system) (right) defense systems. In the Thoeris system, bacteriophage infection activates the host ThsB TIR NADase to produce an isomer of cADPR, which then activates ThsA, leading to NAD^+ depletion and subsequent metabolic crisis and growth arrest. In the Pycsar system, phage infection activates the PycC uridylate cyclase to produce cyclic uridine monophosphate (cUMP). cUMP activates the PycTIR NADase effector to promote abortive infection through NAD^+ depletion. In CBASS, CD-NTase (cGAS/DncV-like nucleotidyltransferase) enzymes generate cyclic oligonucleotides. In the example presented here, c-di-GMP is generated by a member of the CD-NTase family. The ci-di-GMP activates the NADase activity of the TIR-STING protein to cause bacterial growth arrest. Although the genomic location of CD-NTase and TIR-STING within "defense islands" suggests a role in bacterial defense, the phages that activate the TIR-STING system remain unknown.

assembly of signaling complexes. The observation that the animal SARM1 TIR domain expressed an enzymatic function prompted the discovery of NAD^+ hydrolase activity in both plant TIR-only proteins and TNL proteins (*23*, *24*, *95*). Although both animal SARM1 and plant TIR domain proteins trigger cell death after activation, they achieve this through distinct mechanisms. SARM1 possesses high enzymatic activity and rapidly depletes the cell of NAD^+, resulting in energetic catastrophe and death (*22*, *23*, *75*). Plant TIRs are less-active enzymes in biochemical assays, and in vivo TIR activity does not result in severe NAD^+ depletion (*23*). This raises the question of how TIR activity results in cell death if it does not induce energetic catastrophe.

Enhanced disease susceptibility 1 (EDS1) is a conserved immune hub required for plant TIR-dependent cell death and disease resistance phenotypes across plant phylogeny (*96*). Downstream of EDS1 is a small set of ancient and conserved resistance to powdery mildew 8 (RPW8)–like NLRs, called RNL "helper NLRs" that are also required for TIR-mediated cell death and immunity (*97*). RNL helper NLRs transduce TIR-EDS1–dependent signals into direct calcium channel formation and subsequent signaling (*98*, *99*). If plant TIRs functioned through direct NAD^+ depletion, one might expect that they would not require downstream signaling components. Indeed, ectopic expression of SARM1 TIR triggers EDS1-independent cell death in plants (*23*). This is consistent with the hypothesis that strong NADases can kill cells through metabolic dysregulation. Most plant TIRs, by contrast, are EDS1 dependent, even though they are enzymatically active in *eds1* mutants, as measured by the production of v-cADPR (*23*). Thus, plant TIR NADase function is upstream of EDS1 (Fig. 5). A class of TIR proteins is encoded in some plant genomes that lack EDS1 (*100*). It is unclear how these EDS1-independent TIRs function or whether they are producing small-molecule signals.

How, then, does TIR enzymatic activity activate downstream signaling? A pair of studies demonstrated a mechanistic link between TIR enzymatic products and EDS1 complex function (*101*, *102*). Huang *et al.* (*101*) identified a small molecule after coexpressing an active TIR enzyme along with the downstream components EDS1 and PHYTOALEXIN DEFICIENT 4 (PAD4) in insect cells (*101*). EDS1 functions as a heterodimer with one of two structurally related proteins, either PAD4 or SENESCENCE-ASSOCIATED GENE 101 (SAG101). EDS1, PAD4, and SAG101 share a similar domain architecture, with an N-terminal lipase-like domain and a C-terminal domain called the "EP domain" for EDS1 and PAD4. SAG101 also contains an EP domain (Fig. 5). The EDS1-SAG101 crystal structure revealed a cavity formed in the heterodimer, and EP domain residues that face this cavity are required for EDS1 complex function (*103*, *104*). By purifying and crystalizing the TIR-activated EDS1-PAD4 complex, Huang *et al.* found electron density in the EP domain pocket consistent with an ADPR-related molecule that they named phosphoribosyl adenosine diphosphate (pRib-ADP) (*101*). TIR activity induces preferential EDS1-PAD4 heterodimer formation followed by recruitment and activation of downstream ACTIVATED DISEASE RESISTANCE (ADR) RNLs in vivo (*104*–*106*). Using an in vitro assay, Huang *et al.* found that synthetic pRib-ADP and a phosphoribosyl adenosine monophosphate (pRib-AMP) can replace TIR activity that drives the interaction of EDS1-PAD4 heterodimers with downstream ADR RNLs (*101*). The putative TIR catalytic glutamate was required both in vitro to drive EDS1 complex oligomerization and in vivo for the TIR-only protein RBA1 (response to the bacterial type III effector protein HopBA1) to promote pRib-AMP accumulation (*38*, *101*). Thus, this study links TIR enzymatic function mechanistically to EDS1 and downstream events, which likely induce RNLs to form active resistosome-like calcium channels (*98*, *99*) (Fig. 5). In an accompanying research article (*102*), plant TIRs were demonstrated to generate ADP-ribosylated ADPR (di-ADPR) and ADP-ribosylated ATP (ADPr-ATP) (Fig. 3B). These molecules are structurally related to pRib-AMP and pRib-ADP but preferentially activate the EDS1-SAG101 heterodimer via a binding site overlapping that of the pRib molecules (*102*). In vitro, the RPS4 TIR domain was capable of producing all four EDS1 complex signals from NAD^+ and ATP (*102*). These results provide a compelling model that explains how PAD4 or SAG101 form specific EDS1 heterodimeric complexes that selectively recruit and signal through the two subclasses of RNL helper NLRs, ADRs, or NRGs, respectively. This results in distinct outputs, one biased toward cell death and the other toward a transcriptional immune output by means of as-yet-unknown mechanisms (*107*).

The plant protein BdTIR activates ThsA in the Thoeris system (*18*). Thus, plant TIRs can produce a conserved signal (or signals) recognized by downstream signaling molecules across a considerable phylogenetic distance. Similar experiments were performed in the opposite direction by expressing bacterial TIRs in plants (*108*). The human inflammasome perceives intracellular pathogen signals and oligomerizes into a wheel-like disc (*109*, *110*). Using this chassis, Duxbury *et al.* generated a hybrid inflammasome that fused TIR domains to the oligomerization domains of an animal inflammasome (*108*). The hybrid receptor with a plant TIR domain activated cell death in planta in a ligand-specific manner. This death was EDS1 dependent, consistent with activation of the endogenous TIR cell death pathway. Similar fusions with bacterial AbTir, however, failed to trigger cell death but did produce measurable v-cADPR (2′cADPR) (*108*). Duxbury *et al.* concluded that TIR NADase activity and v-cADPR production appear necessary, but not sufficient, for plant cell death (but see section "Plant TIR domains also cleave DNA and RNA" for 2′,3′-cNMP synthetase activity).

One outstanding question is the relevance of plant TIR-produced 2′cADPR (Fig. 3). Is this molecule an intermediate in the production of the ligands that drive specific EDS1 complex formation, does it have distinct functions, or is it immunologically irrelevant enzymatic noise? Could it have a function related to the 3′cADPR produced in the plant cell by phytopathogens via HopAM1? The 2′cADPR molecule was originally described as a "biomarker" for TIR

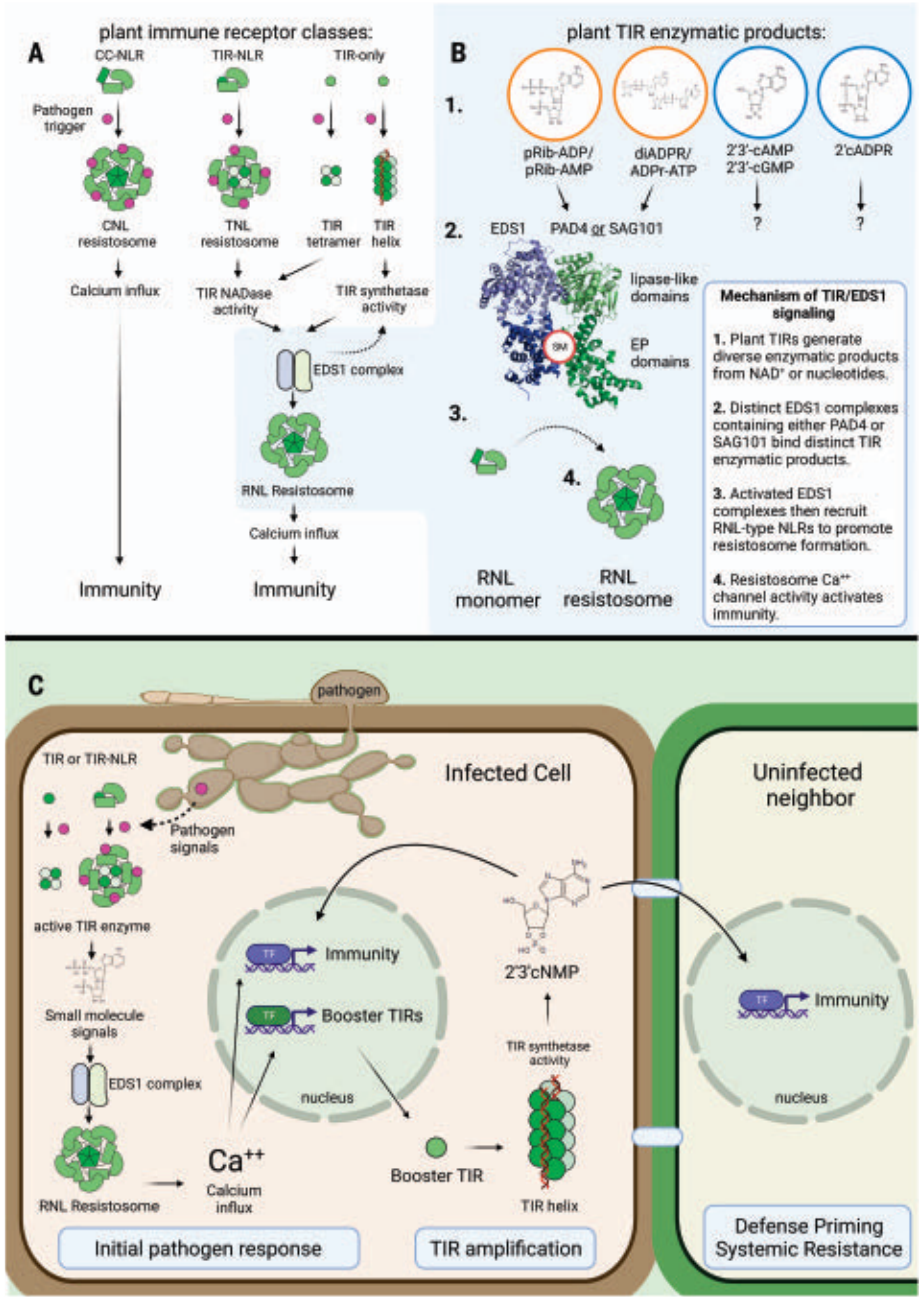


Fig. 5. Intracellular plant immune receptor pathways. (**A**) Plant immune receptors are shown in green. Both CC-NLRs and TIR-NLRs have been shown to bind pathogen triggers (pink circles) and oligomerize into wheel-like resistosomes. For CC-NLRs, the pentameric resistosome generates an ion channel that allows an influx of Ca^{++} and potentially other ions into the cytoplasm. TIR-NLRs form a tetrameric oligomer that activates the NADase activity of the TIR domain. TIR-only proteins can be activated by pathogen triggers to form a presumed NADase tetramer, which produces small-molecule signals (e.g., pRib-ADP) or a helical assembly that expresses both nuclease and 2′3′-cNMP synthetase activity. TNL and TIR-only pathways are dependent on the downstream EDS1 complex and CC_{RPW8}-NLRs (RNL "helper NLRs"). (**B**) The EDS1 complex binds TIR enzymatic products, and this activated complex interacts selectively with ADR or NRG helper RNLs to activate immunity (see inset). SM, small molecules. (**C**) A hypothetical model for a two-tiered TIR immune system. The initial TIR-NLR response is triggered when a TIR-NLR or TIR-only protein recognizes the presence of a pathogen-dependent trigger. This results in production of enzymatic products and activation of the EDS1 complex and RNLs, as shown in (A) and (B). TIR amplification occurs when the initial pathogen response results in the transcription of a large number of "booster TIRs." Booster TIRs amplify the initial response and result in the production of large amounts of 2′3′-cNMP, which then goes on to activate further immune response in both infected and neighboring uninfected cells. TF, transcription factor.

enzymatic function in planta because its presence was correlated with TIR immune function (*23*). It is notable that hydrolysis of the pyrophosphate linkage of 2′cADPR would yield pRib-AMP (Fig. 3C). pRib-ADP has an additional phosphate that is not present in 2′cADPR. Generation of pRib-ADP from 2′cADPR would require either phosphorolysis of the pyrophosphate bond in 2′cADPR (*111*) or phosphorylation of pRib-AMP. 2′cADPR's characteristic 1″-2′ ribose-ribose linkage is also present in (poly)ADP-ribose, a molecule that is induced during plant immune responses (*112*). Cleavage of linear (poly)ADP-ribose could directly produce three of the EDS1-complex signals: pRib-AMP, pRib-ADP, and ADP-ribosylated ADPR (Fig. 3D). It will be interesting to determine whether TIRs can also use polyADP-ribose as a substrate, potentially releasing diADPR through hydrolysis of the 1″-2′ linkage.

Structural basis of TIR-domain enzymatic function

The crystal structures of many animal, plant, and bacterial TIR domains have been determined (*43*, *113*–*117*). A conserved core fold is present in TIR domains that consists of a conserved central five-stranded β sheet (βA to βE), typically with five surrounding helical regions (αA to αE) separated by loops (*10*, *116*). These TIR-domain structures share similarities with the Rossman fold, an ancient motif with alternating β strands and α helices that can bind nucleotides such as NAD^+ (*90*, *118*, *119*). As noted above, mutation of the conserved glutamate residue in TIR domains abolishes enzymatic activity and function in animal, plant, and bacterial systems. Additionally, a conserved aromatic amino acid was identified that is required for generating a cyclized ADPR product (cADPR, 2′cADPR, or 3′cADPR) (*84*). Thus, a shared TIR catalytic mechanism has been conserved throughout evolution.

TIR activation is often driven by oligomerization. Two groups reported cryo–electron

microscopy (cryo-EM) structures for ligand-bound, activated plant TNL receptors (*120*, *121*). The oligomerization state of activated TNL immune receptors bound to their ligands is a tetramer (Fig. 6). The pathogen effector ligand binds primarily to the LRRs and a downstream C-terminal domain and causes the NBS-LRR domains to oligomerize into a ring-like structure. This structure brings four TIR domains into proximity and is reminiscent of the pentameric ZAR1 CNL resistosome (*122*) and several animal inflammasomes. The interactions between these TIR domains in their full-length protein context leverage interfaces predicted from truncated TIR-domain crystal structures to produce a tetramer composed of two pairs of TIRs (*120*, *121*). Within each TIR pair, one of the TIR domains undergoes a conformational change in which the "BB loop" is rearranged to expose the putative catalytic glutamate residue. This transformation upon tetramerization would allow cleavage of NAD^+ and promote production of TIR-derived signal molecules. Cryo-EM studies of SARM1 in its active conformation revealed similar findings, with two SARM1 TIR molecules forming a substrate binding site (*77*). The BB loop is present in this proposed active site for NAD^+ cleavage (Fig. 6), and mutations in the BB loop inhibit NADase activity and axon degeneration (*22*, *24*, *123*). Similar to SARM1, a substrate binding site in bacterial AbTir is formed by two TIR molecules. Mutations in the BB loop also impair enzymatic function (*77*, *84*).

The autoinhibitory regulation of plant NLRs and animal SARM1 share some common features. SARM1 exists as an octamer in a ring shape, with the tandem SAM domains forming the central core of the ring, the N-terminal armadillo repeat (ARM) domains wrapping around the ring, and the TIR domains lodged between the ARM domains on the outside of the ring (*77*, *114*, *115*, *124*–*127*). The N-terminal autoinhibitory ARM domain in SARM1 interacts with the TIR domain (*77*, *114*, *115*, *123*–*126*). This is analogous to the NBS and LRR domains in plant NLRs (*7*). After activation, the eight SARM1 TIR domains move above the plane of the ARM-SAM ring and reorganize in a two-by-four grid (*77*). The organization of the SARM1 TIRs is notably similar to that of the plant tetrameric TIR, using the same interfaces and overall geometry (*77*).

During activation of SARM1, both NAD^+ and NMN compete for an allosteric binding site within the autoinhibitory N-terminal domain of SARM1, with the ratio of bound NMN to NAD^+ dictating the activation state of the enzyme (*114*, *115*, *125*, *126*, *128*). Higher NMN-to-NAD^+ ratios result in an activated enzyme that consumes NAD^+ and promotes axon death, whereas low ratios keep the enzyme in the off state (*114*). Mutations in the NMN-NAD^+ binding pocket that disrupt NMN binding block both activation of NAD^+ hydrolase activity and injury-induced axon degeneration (*114*). Conversely, hypermorphic mutations in the NMN binding pocket produce an enhanced or constitutively active form of the enzyme (*114*). These structural studies of SARM1 clarify the activation mechanism of the TIR enzymatic activity and provide a blueprint for how other proteins with TIR-domain NAD^+ hydrolases may be regulated.

Although plant TNLs and animal SARM1 have convergent mechanisms for TIR activation, not all TIRs follow their example as closely. The bacterial TIR-STING (TIR–stimulator of interferon genes) sensor is present in both prokaryotes and basal eukaryotes (*25*). TIR domains of TIR-STING proteins use a proline-rich loop to mediate TIR-TIR interactions but rely on the same conserved catalytic glutamate for NADase activity (Fig. 6) (*25*). The binding of cyclic-di-guanine monophosphate (c-di-GMP) to the STING domain of TIR-STING proteins drives oligomerization into filamentous

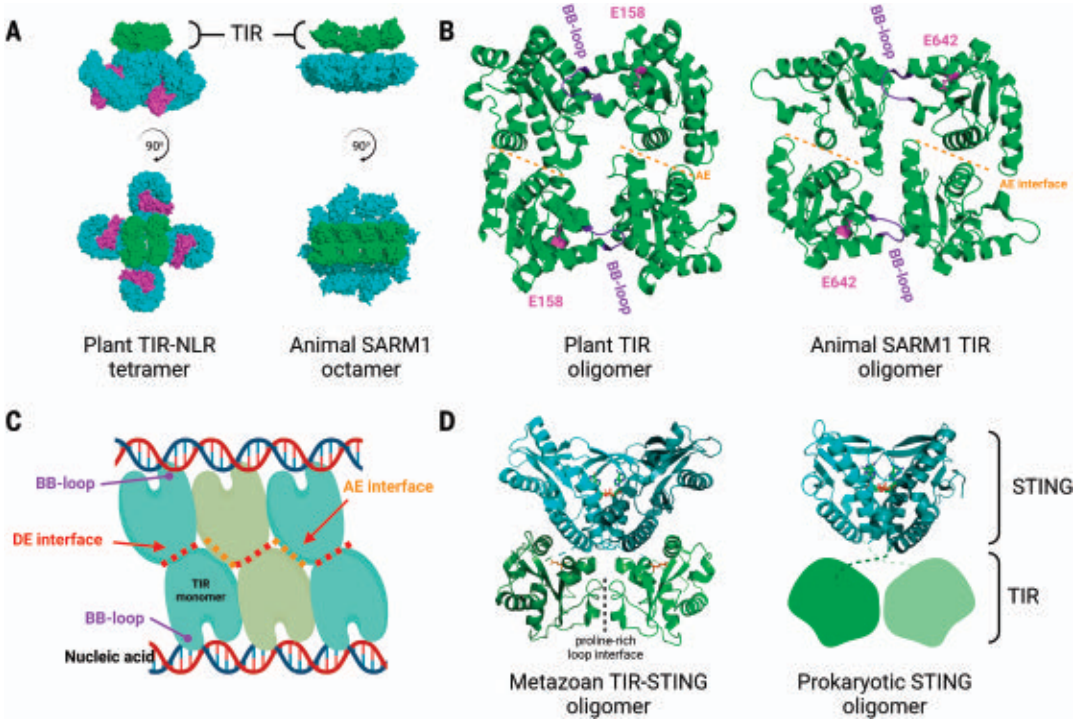


Fig. 6. TIR protein oligomerization strategies. (**A**) Activated plant TIR-NLR immune receptors and animal SARM1 function as oligomeric proteins. TIR domains are shown in green, the oligomeric chassis is in teal, and, for the plant complex, the bacterial ligand is in magenta. The oligomerized plant TIR complex is tetrameric [RPP1—Protein Data Bank (PDB) ID 7CRC—is shown; also see ROQ1 (PDB ID 7JLX)], and the animal SARM1 is octameric (PDB IDs 7NAK and 7NAL). (**B**) Convergent oligomerization strategies shown in (A) result in similar TIR interfaces for plant TIRs (PDB ID 7CRC is shown) and SARM1. Putative catalytic glutamates are shown in magenta. The BB loop is shown in purple, and the interaction interface mediated by the αA and αE helices (AE) is shown in orange. Graphics in (A) and (B) are adapted from Shi *et al*. (*77*). (**C**) Schematic cartoon of a plant TIR–nucleic acid helical oligomer. In contrast to the tetrameric TIR-NLR structure, helical TIRs are interconnected with repeating AE (orange) and DE (red) interfaces. The DE interaction interface is mediated by the αD and αE helices. The BB loop interacts with the minor groove of DNA. The cartoon is adapted from Yu *et al*. (*27*). (**D**) Both metazoans and prokaryotes have TIR NADases that are regulated by C-terminal STING domains. Oligomerization of TIR-STING proteins is controlled by cyclic-dinucleotide (sticks) binding to the STING domain (teal). The two monomers cross perpendicular to each other, indicated by dark and light shading. Green TIR domains associate through a distinctive proline-rich loop interaction surface (black dashed line). Putative catalytic glutamate is shown in orange. The prokaryotic TIR domain was not crystalized but is potentially similar to the metazoan structure. The metazoan TIR-STING structure is PDB ID 6WT7, and the prokaryotic STING structure is PDB ID 6WT4. The graphic in (D) is adapted from Morehouse *et al*. (*25*).

structures, and mutations that disrupt either c-di-GMP binding or oligomerization into filaments disrupt NAD^+ cleavage (*25*). The formation of filamentous structures was also observed upon binding of small molecules to additional bacterial TIR NADases (*84*, *129*). The function of these higher-order structures remains to be elucidated but is reminiscent of prion-like protein activation seen in other immune defense proteins (*130*, *131*). Thus, there are diverse, independently evolved strategies to oligomerize and/or activate TIR NADase activity.

Plant TIR domains also cleave DNA and RNA

Plant TIR enzymatic activity that produces small-molecule signals to activate either of the EDS1 RNL complexes is essential for function. A second enzymatic function was described for plant TIRs (*27*). Yu *et al.* found that both TIR-only proteins and TIR domains derived from TNL proteins could degrade nucleotides, DNA, and RNA in vitro. Plant TIR domains produce cyclic nucleotide monophosphates (cNMPs) such as 2′,3′-cAMP and 2′,3′-cGMP, molecules that are stress signals in plants and animals (*132*, *133*).

Whereas the first plant TNL cryo-EM structures revealed a tetrameric resistosome, new cryo-EM structures show that a plant TIR domain (lacking NBS-LRR domains) is capable of oligomerizing into helical structures (*27*). The purified L7 TIR domain is a C-terminal truncation that lacks the NBS-LRR domains that are found in the full-length L7 TNL protein. The L7 TIR domain structure revealed two protofilaments that twist into a superhelix that also appears to contain nucleic acid (Fig. 6). This structure supports the hypothesis that plant TIRs can also act as nucleases and defines basic arginine and lysine residues that likely contact the nucleic acid. Indeed, the mutation of these residues completely blocks 2′,3′-cNMP synthetase activity and TIR-triggered cell death but only partially impairs NADase function. Thus, NADase and synthetase activity are separable, and partial NADase activity alone is insufficient for immune function as assayed by cell death. Although the structural details of catalysis are unclear, TIR nuclease activity and 2′,3′ synthetase activity both share the requirement for the putative catalytic glutamic acid with NADase activity, indicating functional overlap.

In planta experiments that genetically reduce levels of 2′,3′-cNMP also blocked TIR cell death, consistent with an immune signaling function for these cyclic nucleotide products. Is 2′,3′-cNMP another signal that activates EDS1? Or does this product have other signaling functions? Accumulation of 2′,3′-cNMP in planta requires the EDS1 complex, which would place its generation downstream of EDS1, at odds with data showing that TIR enzymes act upstream of EDS1 (*27*). To explain this, Yu *et al.* suggest a feedback loop in which the putative 2′,3′-cNMP EDS1-activating signal must be amplified in an EDS1-dependent manner (Fig. 5) (*27*). Yu *et al.* also found that bacterial AbTir does not have 2′,3′-cNMP synthetase activity. The lack of AbTir synthetase activity could potentially explain why this 2′cADPR-producing TIR did not activate the EDS1-dependent plant cell death pathway as a chimeric fusion to the NLRC4 chassis, as described above (*27*, *108*). The authors conclude that plant TIRs can exist in two distinct oligomeric complexes: as a tetramer with NADase activity and as a helical synthetase. Many details of this dual-specificity enzymatic system remain to be clarified. Are full-length TIR-NBS-LRR proteins capable of forming the helical synthetase given potential steric issues? Do alternatively spliced TIRs related to the truncated TIR-only isoform generate 2′3′-cNMP? Is there selective catalysis to produce both the EDS1-complex signals and 2′,3′-cNMP products? And what are the relevant in vivo substrates? Immune activation leads to transcriptional activation of many TIR domain–encoding genes, and overexpression of these genes can cause cell death (*39*, *134*–*136*). Thus, one model is that a primary TNL or TIR-only receptor is activated by a pathogen-encoded ligand to promote oligomerization that favors NAD^+ hydrolase activity; the consequent EDS1 complex–dependent transcriptional output generally raises TIR domain levels through new transcription (Fig. 5C). These "booster TIRs," in turn, drive helical oligomerization that favors 2′3′-cNMP synthetase activity to stimulate defense responses, perhaps in different subcellular compartments and/or in cells directly neighboring the infection site (Fig. 5C).

Conclusion

The study of TIR domain proteins as an enzyme family has transformed our understanding of this evolutionarily ancient signaling domain. The discovery of the SARM1 NADase has directed therapeutic approaches in neurodegeneration (*137*–*140*). Constitutively hyperactive SARM1 NADase variants were found in ALS patients (*141*, *142*), suggesting that SARM1 alleles may contribute to disease risk in ALS. Hence, inhibiting SARM1 could alleviate certain neurodegenerative diseases. The role of SARM1 and NAD^+ metabolism during development and aging is also emerging (*143*–*147*). Future studies will help strengthen this link, with implications for neurodevelopmental genetic disorders.

Do TLRs and other animal TIR domains besides SARM1 truly lack enzymatic function? A few published reports suggest this (*22*, *24*, *81*). Yet some of these proteins possess the conserved glutamic acid residue necessary for NAD^+ cleavage in SARM1, plant TIRs, and bacterial TIRs. Structural studies indicate these glutamate residues are in suboptimal positions to perform the necessary catalytic functions (*24*, *81*). It is possible that mammalian TIR domains retain enzymatic activities but that their cognate substrates have not yet been identified. Prior biochemical experiments with purified proteins may have failed to recapitulate the true biologically active conformation of these mammalian proteins and hence missed their enzymatic activity.

The discovery of new cADPR isomers generated by plant and bacterial TIR proteins (Fig. 3) is another advance emerging from the work on TIR enzymatic functions. We now know that there are at least two distinct linkages: 2′cADPR and 3′cADPR. These small molecules can bind to and activate other effector enzymes in bacterial antiphage defense programs. Moreover, this ability to activate effector proteins to trigger bacterial suicide can be leveraged to develop a class of antibiotics that will target bacterial TIR proteins during infections. The structural similarities of the cADPR isomers also raise the possibility that these molecules, whether in their native biological context or whether repurposed as drugs, can function as competitive inhibitors of immune pathway proteins to subvert host immune responses. It also remains to be determined if these cADPR isomers possess calcium-mobilizing effects.

Understanding the role of TIR-derived small-molecule products in the plant immune system should clarify gaps in our understanding of plant immunity. Discovery of 2′-linked ribose-ribose small molecules as the connection between TIR enzymatic activity and EDS1 complex association with downstream RNLs is a landmark discovery and may provide new natural chemicals to control disease resistance. Yu *et al.* (*27*) propose that the protein structures produced by TIR-only and TNL proteins have distinct enzymatic activities. How and where these protein structures and activities are occurring in vivo remains an open question for which we have proposed a speculative model (Fig. 5C). Similarly, it is unknown whether full-length TNL proteins (and alternatively spliced products) form both the tetramer NADase and the helical 2′,3′-cNMP synthetase. Equally curious is the potential for separation of functions encoded by full-length TNLs and TIR domain–only proteins like RBA1.

The revelation that TIR proteins are performing similar functions in cell death and disease resistance across kingdoms has had an energizing effect across many fields. Discoveries in animals, plants, and prokaryotes are generating hypotheses and driving research and applications across kingdoms as we learn more about the common and distinctive features of members of this broad and ancient protein family.

REFERENCES AND NOTES

1. S. Akira, S. Uematsu, O. Takeuchi, Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006). doi: 10.1016/j.cell.2006.02.015; pmid: 16497588
2. P. C. Ronald, B. Beutler, Plant and animal sensors of conserved microbial signatures. *Science* **330**, 1061–1064 (2010). doi: 10.1126/science.1189468; pmid: 21097929
3. R. Medzhitov, P. Preston-Hurlburt, C. A. Janeway Jr., A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997). doi: 10.1038/41131; pmid: 9237759
4. K. A. Fitzgerald, J. C. Kagan, Toll-like receptors and the control of immunity. *Cell* **180**, 1044–1066 (2020). doi: 10.1016/j.cell.2020.02.041; pmid: 32164908
5. B. Everts *et al*., TLR-driven early glycolytic reprogramming via the kinases TBK1-IKKε supports the anabolic demands of dendritic cell activation. *Nat. Immunol.* **15**, 323–332 (2014). doi: 10.1038/ni.2833; pmid: 24562310
6. Y. Tan, J. C. Kagan, Innate immune signaling organelles display natural and programmable signaling flexibility. *Cell* **177**, 384–398.e11 (2019). doi: 10.1016/j.cell.2019.01.039; pmid: 30853218
7. J. D. G. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016). doi: 10.1126/science.aaf6395; pmid: 27934708
8. A.-L. Van de Weyer *et al*., A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* **178**, 1260–1272.e14 (2019). doi: 10.1016/j.cell.2019.07.038; pmid: 31442410
9. J. D. G. Jones, J. L. Dangl, The plant immune system. *Nature* **444**, 323–329 (2006). doi: 10.1038/nature05286; pmid: 17108957
10. V. Y. Toshchakov, A. F. Neuwald, A survey of TIR domain sequence and structure divergence. *Immunogenetics* **72**, 181–203 (2020). doi: 10.1007/s00251-020-01157-7; pmid: 32002590
11. A. M. Spear, N. J. Loman, H. S. Atkins, M. J. Pallen, Microbial TIR domains: Not necessarily agents of subversion? *Trends Microbiol.* **17**, 393–398 (2009). doi: 10.1016/j.tim.2009.06.005; pmid: 19716705
12. R. M. Newman, P. Salunkhe, A. Godzik, J. C. Reed, Identification and characterization of a novel bacterial virulence factor that shares homology with mammalian Toll/interleukin-1 receptor family proteins. *Infect. Immun.* **74**, 594–601 (2006). doi: 10.1128/IAI.74.1.594-601.2006; pmid: 16369016
13. C. Cirl *et al*., Subversion of Toll-like receptor signaling by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins. *Nat. Med.* **14**, 399–406 (2008). doi: 10.1038/nm1734; pmid: 18327267
14. Q. Zhang, C. M. Zmasek, X. Cai, A. Godzik, TIR domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. *Dev. Comp. Immunol.* **35**, 461–468 (2011). doi: 10.1016/j.dci.2010.11.013; pmid: 21110998
15. K. Essuman *et al*., TIR domain proteins are an ancient family of NAD$^+$-consuming enzymes. *Curr. Biol.* **28**, 421–430.e4 (2018). doi: 10.1016/j.cub.2017.12.024; pmid: 29395922
16. S. Doron *et al*., Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). doi: 10.1126/science.aar4120; pmid: 29371424
17. J. S. Weagley *et al*., Products of gut microbial Toll/interleukin-1 receptor domain NADase activities in gnotobiotic mice and Bangladeshi children with malnutrition. *Cell Rep.* **39**, 110738 (2022). doi: 10.1016/j.celrep.2022.110738; pmid: 35476981
18. G. Ofir *et al*., Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116–120 (2021). doi: 10.1038/s41586-021-04098-7; pmid: 34853457
19. L. A. J. O'Neill, D. Golenbock, A. G. Bowie, The history of Toll-like receptors – redefining innate immunity. *Nat. Rev. Immunol.* **13**, 453–460 (2013). doi: 10.1038/nri3446; pmid: 23681101
20. J. D. Nanson, B. Kobe, T. Ve, Death, TIR, and RHIM: Self-assembling domains involved in innate immunity and cell-death signaling. *J. Leukoc. Biol.* **105**, 363–375 (2019). doi: 10.1002/JLB.MR0318-123R; pmid: 30517972
21. S. Nimma, T. Ve, S. J. Williams, B. Kobe, Towards the structure of the TIR-domain signalosome. *Curr. Opin. Struct. Biol.* **43**, 122–130 (2017). doi: 10.1016/j.sbi.2016.12.014; pmid: 28092811
22. K. Essuman *et al*., The SARM1 Toll/interleukin-1 receptor domain possesses intrinsic NAD$^+$ cleavage activity that promotes pathological axonal degeneration. *Neuron* **93**, 1334–1343.e5 (2017). doi: 10.1016/j.neuron.2017.02.022; pmid: 28334607
23. L. Wan *et al*., TIR domains of plant immune receptors are NAD$^+$-cleaving enzymes that promote cell death. *Science* **365**, 799–803 (2019). doi: 10.1126/science.aax1771; pmid: 31439793
24. S. Horsefield *et al*., NAD$^+$ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019). doi: 10.1126/science.aax1911; pmid: 31439792
25. B. R. Morehouse *et al*., STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020). doi: 10.1038/s41586-020-2719-5; pmid: 32877915
26. Z. Y. Zhao *et al*., A cell-permeant mimetic of NMN activates SARM1 to produce cyclic ADP-ribose and induce non-apoptotic cell death. *iScience* **15**, 452–466 (2019). doi: 10.1016/j.isci.2019.05.001; pmid: 31128467
27. D. Yu *et al*., TIR domains of plant immune receptors are 2′,3′-cAMP/cGMP synthetases mediating cell death. *Cell* **185**, 2370–2386.e18 (2022). doi: 10.1016/j.cell.2022.04.032; pmid: 35597242
28. J. E. Sims *et al*., cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science* **241**, 585–589 (1988). doi: 10.1126/science.2969618; pmid: 2969618
29. N. J. Gay, F. J. Keith, *Drosophila* Toll and IL-1 receptor. *Nature* **351**, 355–356 (1991). doi: 10.1038/351355b0; pmid: 1851964
30. D. S. Schneider, K. L. Hudson, T. Y. Lin, K. V. Anderson, Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo. *Genes Dev.* **5**, 797–807 (1991). doi: 10.1101/gad.5.5.797; pmid: 1827421
31. K. V. Anderson, G. Jürgens, C. Nüsslein-Volhard, Establishment of dorsal-ventral polarity in the Drosophila embryo: Genetic studies on the role of the *Toll* gene product. *Cell* **42**, 779–789 (1985). doi: 10.1016/0092-8674(85)90274-0; pmid: 3931918
32. R. Steward, *Dorsal*, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, c-*rel*. *Science* **238**, 692–694 (1987). doi: 10.1126/science.3118464; pmid: 3118464
33. E. Stylianou *et al*., Interleukin 1 induces NF-κB through its type I but not its type II receptor in lymphocytes. *J. Biol. Chem.* **267**, 15836–15841 (1992). doi: 10.1016/S0021-9258(19)49611-7; pmid: 1386364
34. S. Whitham *et al*., The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115 (1994). doi: 10.1016/0092-8674(94)90283-6; pmid: 7923359
35. Y. T. Ip *et al*., *Dif*, a *dorsal*-related gene that mediates an immune response in *Drosophila*. *Cell* **75**, 753–763 (1993). doi: 10.1016/0092-8674(93)90495-C; pmid: 8242747
36. B. Lemaitre, E. Nicolas, L. Michaut, J. M. Reichhart, J. A. Hoffmann, The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996). doi: 10.1016/S0092-8674(00)80172-5; pmid: 8808632
37. S. Rutschmann *et al*., The Rel protein DIF mediates the antifungal but not the antibacterial host defense in *Drosophila*. *Immunity* **12**, 569–580 (2000). doi: 10.1016/S1074-7613(00)80208-3; pmid: 10843389
38. M. T. Nishimura *et al*., TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2053–E2062 (2017). doi: 10.1073/pnas.1620973114; pmid: 28137883
39. R. S. Nandety *et al*., The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* **162**, 1459–1472 (2013). doi: 10.1104/pp.113.219162; pmid: 23735504
40. M. E. Santamaría *et al*., An Arabidopsis TIR-lectin two-domain protein confers defense properties against *Tetranychus urticae*. *Plant Physiol.* **179**, 1298–1314 (2019). doi: 10.1104/pp.18.00951; pmid: 30765478
41. P. R. Vajjhala, T. Ve, A. Bentham, K. J. Stacey, B. Kobe, The molecular mechanisms of signaling by cooperative assembly formation in innate immunity pathways. *Mol. Immunol.* **86**, 23–37 (2017). doi: 10.1016/j.molimm.2017.02.012; pmid: 28249680
42. J. C. Kagan, V. G. Magupalli, H. Wu, SMOCs: Supramolecular organizing centres that control innate immunity. *Nat. Rev. Immunol.* **14**, 821–826 (2014). doi: 10.1038/nri3757; pmid: 25359439
43. T. Ve *et al*., Structural basis of TIR-domain-assembly formation in MAL- and MyD88-dependent TLR4 signaling. *Nat. Struct. Mol. Biol.* **24**, 743–751 (2017). doi: 10.1038/nsmb.3444; pmid: 28759049
44. S. Nimma *et al*., Structural evolution of TIR-domain signalosomes. *Front. Immunol.* **12**, 784484 (2021). doi: 10.3389/fimmu.2021.784484; pmid: 34868065
45. C. Garlanda, C. A. Dinarello, A. Mantovani, The interleukin-1 family: Back to the future. *Immunity* **39**, 1003–1018 (2013). doi: 10.1016/j.immuni.2013.11.010; pmid: 24332029
46. C. A. Dinarello, The IL-1 family of cytokines and receptors in rheumatic diseases. *Nat. Rev. Rheumatol.* **15**, 612–632 (2019). doi: 10.1038/s41584-019-0277-8; pmid: 31515542
47. J. E. Sims, D. E. Smith, The IL-1 family: Regulators of immunity. *Nat. Rev. Immunol.* **10**, 89–102 (2010). doi: 10.1038/nri2691; pmid: 20081871
48. M. Carty *et al*., The human adaptor SARM negatively regulates adaptor protein TRIF-dependent Toll-like receptor signaling. *Nat. Immunol.* **7**, 1074–1081 (2006). doi: 10.1038/ni1382; pmid: 16964262
49. C. Couillault *et al*., TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. *Nat. Immunol.* **5**, 488–494 (2004). doi: 10.1038/ni1060; pmid: 15048112
50. C.-F. Chuang, C. I. Bargmann, A Toll-interleukin 1 repeat protein at the synapse specifies asymmetric odorant receptor expression via ASK1 MAPKKK signaling. *Genes Dev.* **19**, 270–281 (2005). doi: 10.1101/gad.1276505; pmid: 15625192
51. J. E. Irazoqui, J. M. Urbach, F. M. Ausubel, Evolution of host innate defence: Insights from *Caenorhabditis elegans* and primitive invertebrates. *Nat. Rev. Immunol.* **10**, 47–58 (2010). doi: 10.1038/nri2689; pmid: 20029447
52. J. Gerdts, D. W. Summers, J. Milbrandt, A. DiAntonio, Axon self-destruction: New links among SARM1, MAPKs, and NAD$^+$ metabolism. *Neuron* **89**, 449–460 (2016). doi: 10.1016/j.neuron.2015.12.023; pmid: 26844829
53. M. D. Figley, A. DiAntonio, The SARM1 axon degeneration pathway: Control of the NAD$^+$ metabolome regulates axon survival in health and disease. *Curr. Opin. Neurobiol.* **63**, 59–66 (2020). doi: 10.1016/j.conb.2020.02.012; pmid: 32311648
54. A. Waller, Experiments on the section of the glosso-pharyngeal and hypoglossal nerves of the frog, and observations of the alterations produced thereby in the structure of their primitive fibres. *Edinb. Med. Surg. J.* **76**, 369–376 (1851). pmid: 30332247
55. E. R. Lunn, V. H. Perry, M. C. Brown, H. Rosen, S. Gordon, Absence of Wallerian degeneration does not hinder regeneration in peripheral nerve. *Eur. J. Neurosci.* **1**, 27–33 (1989). doi: 10.1111/j.1460-9568.1989.tb00771.x; pmid: 12106171
56. T. G. A. Mack *et al*., Wallerian degeneration of injured axons and synapses is delayed by a *Ube4b/Nmnat* chimeric gene. *Nat. Neurosci.* **4**, 1199–1206 (2001). doi: 10.1038/nn770; pmid: 11770485
57. L. Conforti *et al*., A *Ufd2/D4Cole1e* chimeric protein and overexpression of *Rbp7* in the slow Wallerian degeneration (*Wld*^{S}) mouse. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11377–11382 (2000). doi: 10.1073/pnas.97.21.11377; pmid: 11027338
58. T. Araki, Y. Sasaki, J. Milbrandt, Increased nuclear NAD biosynthesis and SIRT1 activation prevent axonal degeneration. *Science* **305**, 1010–1013 (2004). doi: 10.1126/science.1098014; pmid: 15310905
59. B. R. Miller *et al*., A dual leucine kinase-dependent axon self-destruction program promotes Wallerian degeneration. *Nat. Neurosci.* **12**, 387–389 (2009). doi: 10.1038/nn.2290; pmid: 19287387
60. J. Yang *et al*., Pathological axonal death through a MAPK cascade that triggers a local energy deficit. *Cell* **160**, 161–176 (2015). doi: 10.1016/j.cell.2014.11.053; pmid: 25594179
61. J. M. Osterloh *et al*., dSarm/Sarm1 is required for activation of an injury-induced axon death pathway. *Science* **337**, 481–484 (2012). doi: 10.1126/science.1223899; pmid: 22678360
62. J. Gerdts, D. W. Summers, Y. Sasaki, A. DiAntonio, J. Milbrandt, Sarm1-mediated axon degeneration requires both SAM and TIR interactions. *J. Neurosci.* **33**, 13569–13580 (2013). doi: 10.1523/JNEUROSCI.1197-13.2013; pmid: 23946415
63. M. P. Coleman, A. Höke, Programmed axon degeneration: From mouse to mechanism to medicine. *Nat. Rev. Neurosci.*

21, 183–196 (2020). doi: 10.1038/s41583-020-0269-3; pmid: 32152523
64. N. Henninger *et al*., Attenuated traumatic axonal injury and improved functional outcome after traumatic brain injury in mice lacking *Sarm1*. *Brain* **139**, 1094–1105 (2016). doi: 10.1093/brain/aww001; pmid: 26912636
65. S. Geisler *et al*., Prevention of vincristine-induced peripheral neuropathy by genetic deletion of SARM1 in mice. *Brain* **139**, 3092–3108 (2016). doi: 10.1093/brain/aww251; pmid: 27797810
66. M. A. White *et al*., *Sarm1* deletion suppresses TDP-43-linked motor neuron degeneration and cortical spine loss. *Acta Neuropathol. Commun.* **7**, 166 (2019). doi: 10.1186/s40478-019-0800-9; pmid: 31661035
67. Y. Cheng *et al*., *Sarm1* gene deficiency attenuates diabetic peripheral neuropathy in mice. *Diabetes* **68**, 2120–2130 (2019). doi: 10.2337/db18-1233; pmid: 31439642
68. E. Turkiew, D. Falconer, N. Reed, A. Höke, Deletion of Sarm1 gene is neuroprotective in two models of peripheral neuropathy. *J. Peripher. Nerv. Syst.* **22**, 162–171 (2017). doi: 10.1111/jns.12219; pmid: 28485482
69. Y. Sasaki *et al*., SARM1 depletion rescues NMNAT1-dependent photoreceptor cell death and retinal degeneration. *eLife* **9**, e62027 (2020). doi: 10.7554/eLife.62027; pmid: 33107823
70. K. W. Ko, J. Milbrandt, A. DiAntonio, SARM1 acts downstream of neuroinflammatory and necroptotic signaling to induce axon degeneration. *J. Cell Biol.* **219**, e201912047 (2020). doi: 10.1083/jcb.201912047; pmid: 32609299
71. Y.-H. Chen, Y. Sasaki, A. DiAntonio, J. Milbrandt, SARM1 is required in human derived sensory neurons for injury-induced and neurotoxic axon degeneration. *Exp. Neurol.* **339**, 113636 (2021). doi: 10.1016/j.expneurol.2021.113636; pmid: 33548217
72. K. J. Szretter *et al*., The immune adaptor molecule SARM modulates tumor necrosis factor alpha production and microglia activation in the brainstem and restricts West Nile virus pathogenesis. *J. Virol.* **83**, 9329–9338 (2009). doi: 10.1128/JVI.00836-09; pmid: 19587044
73. C. G. Doran *et al*., CRISPR/Cas9-mediated SARM1 knockout and epitope-tagged mice reveal that SARM1 does not regulate nuclear transcription, but is expressed in macrophages. *J. Biol. Chem.* **297**, 101417 (2021). doi: 10.1016/j.jbc.2021.101417; pmid: 34793837
74. R. Krauss, T. Bosanac, R. Devraj, T. Engber, R. O. Hughes, Axons matter: The promise of treating neurodegenerative disorders by targeting SARM1-mediated axonal degeneration. *Trends Pharmacol. Sci.* **41**, 281–293 (2020). doi: 10.1016/j.tips.2020.01.006; pmid: 32107050
75. J. Gerdts, E. J. Brace, Y. Sasaki, A. DiAntonio, J. Milbrandt, SARM1 activation triggers axon degeneration locally via NAD^+ destruction. *Science* **348**, 453–457 (2015). doi: 10.1126/science.1258366; pmid: 25908823
76. C. Angeletti *et al*., SARM1 is a multi-functional NAD(P)ase with prominent base exchange activity, all regulated bymultiple physiologically relevant NAD metabolites. *iScience* **25**, 103812 (2022). doi: 10.1016/j.isci.2022.103812; pmid: 35198877
77. Y. Shi *et al*., Structural basis of SARM1 activation, substrate recognition, and inhibition by small molecules. *Mol. Cell* **82**, 1643–1659.e10 (2022). doi: 10.1016/j.molcel.2022.03.007; pmid: 35334231
78. K. W. Ko, L. Devault, Y. Sasaki, J. Milbrandt, A. DiAntonio, Live imaging reveals the cellular events downstream of SARM1 activation. *eLife* **10**, e71148 (2021). doi: 10.7554/eLife.71148; pmid: 34779400
79. Y. Li *et al*., Sarm1 activation produces cADPR to increase intra-axonal Ca^{++} and promote axon degeneration in PIPN. *J. Cell Biol.* **221**, e202106080 (2022). doi: 10.1083/jcb.202106080; pmid: 34935867
80. Y. Sasaki, T. Nakagawa, X. Mao, A. DiAntonio, J. Milbrandt, NMNAT1 inhibits axon degeneration via blockade of SARM1-mediated NAD^+ depletion. *eLife* **5**, e19749 (2016). doi: 10.7554/eLife.19749; pmid: 27735788
81. S. Nimma *et al*., Crystal structure of the Toll/interleukin-1 receptor (TIR) domain of IL-1R10 provides structural insights into TIR domain signalling. *FEBS Lett.* **596**, 886–897 (2022). doi: 10.1002/1873-3468.14288; pmid: 35038778
82. L. A. J. O'Neill, R. J. Kishton, J. Rathmell, A guide to immunometabolism for immunologists. *Nat. Rev. Immunol.* **16**, 553–565 (2016). doi: 10.1038/nri.2016.70; pmid: 27396447
83. C. Cantó, K. J. Menzies, J. Auwerx, NAD^+ metabolism and the control of energy homeostasis: A balancing act between mitochondria and the nucleus. *Cell Metab.* **22**, 31–53 (2015). doi: 10.1016/j.cmet.2015.05.023; pmid: 26118927
84. M. K. Manik *et al*., Chemical structures of cyclic ADP ribose (cADPR) isomers and the molecular basis of their production and signaling. bioRxiv 2022.05.07.491051 [Preprint] (2022); doi: 10.1101/2022.05.07.491051
85. S. Eastman *et al*., A phytobacterial TIR domain effector manipulates NAD^+ to promote virulence. *New Phytol.* **233**, 890–904 (2022). doi: 10.1111/nph.17805; pmid: 34657283
86. A. Leavitt *et al*., Viruses inhibit TIR gcADPR signaling to overcome bacterial defense. bioRxiv 2022.05.03.490397 [Preprint] (2022). doi: 10.1101/2022.05.03.490397
87. A. Waldhuber *et al*., A comparative analysis of the mechanism of Toll-like receptor-disruption by TIR-containing protein C from uropathogenic *Escherichia coli*. *Pathogens* **5**, 25 (2016). doi: 10.3390/pathogens5010025; pmid: 26938564
88. F. Askarian *et al*., A *Staphylococcus aureus* TIR domain protein virulence factor blocks TLR2-mediated NF-κB signaling. *J. Innate Immun.* **6**, 485–498 (2014). doi: 10.1159/000357618; pmid: 24481289
89. J. M. Coronas-Serna *et al*., The TIR-domain containing effectors BtpA and BtpB from *Brucella abortus* impact NAD metabolism. *PLOS Pathog.* **16**, e1007979 (2020). doi: 10.1371/journal.ppat.1007979; pmid: 32298382
90. D. Ka, H. Oh, E. Park, J.-H. Kim, E. Bae, Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD^+ degradation. *Nat. Commun.* **11**, 2816 (2020). doi: 10.1038/s41467-020-16703-w; pmid: 32499527
91. A. Lopatina, N. Tal, R. Sorek, Abortive infection: Bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020). doi: 10.1146/annurev-virology-011620-040628; pmid: 32559405
92. N. Tal, R. Sorek, SnapShot: Bacterial immunity. *Cell* **185**, 578–578.e1 (2022). doi: 10.1016/j.cell.2021.12.029; pmid: 35120666
93. N. Tal *et al*., Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728–5739.e16 (2021). doi: 10.1016/j.cell.2021.09.031; pmid: 34644530
94. B. Koopal *et al*., Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* **185**, 1471–1486.e19 (2022). doi: 10.1016/j.cell.2022.03.012; pmid: 35381200
95. A. M. Bayless, M. T. Nishimura, Enzymatic functions for Toll/interleukin-1 receptor domain proteins in the plant immune system. *Front. Genet.* **11**, 539 (2020). doi: 10.3389/fgene.2020.00539; pmid: 32582284
96. D. Lapin, D. D. Bhandari, J. E. Parker, Origins and immunity networking functions of EDS1 family proteins. *Annu. Rev. Phytopathol.* **58**, 253–276 (2020). doi: 10.1146/annurev-phyto-010820-012840; pmid: 32396762
97. L. M. Jubic, S. Saile, O. J. Furzer, F. El Kasmi, J. L. Dangl, Help wanted: Helper NLRs and plant immune responses. *Curr. Opin. Plant Biol.* **50**, 82–94 (2019). doi: 10.1016/j.pbi.2019.03.013; pmid: 31063902
98. G. Bi *et al*., The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell* **184**, 3528–3541.e12 (2021). doi: 10.1016/j.cell.2021.05.003; pmid: 33984278
99. P. Jacob *et al*., Plant "helper" immune receptors are Ca^{2+}-permeable nonselective cation channels. *Science* **373**, 420–425 (2021). doi: 10.1126/science.abg7917; pmid: 34140391
100. O. Johanndrees *et al*., Differential EDS1 requirement for cell death activities of plant TIR-domain proteins. bioRxiv 2021.11.29.470438 [Preprint] (2021). doi: 10.1101/2021.11.29.470438
101. S. Huang *et al*., Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity. *Science* **377**, eabq3297 (2022). doi: 10.1126/science.abq3297
102. A. Jia *et al*., TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity. *Science* **377**, eabq8180 (2022). doi: 10.1126/science.abq8180
103. D. D. Bhandari *et al*., An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in Arabidopsis. *Nat. Commun.* **10**, 772 (2019). doi: 10.1038/s41467-019-08783-0; pmid: 30770836
104. J. A. Dongus *et al*., Cavity surface residues of PAD4 and SAG101 contribute to EDS1 dimer signaling specificity in plant immunity. *Plant J.* **110**, 1415–1432 (2022). doi: 10.1111/tpj.15747; pmid: 35324052
105. X. Sun *et al*., Pathogen effector recognition-dependent association of NRG1 with EDS1 and SAG101 in TNL receptor immunity. *Nat. Commun.* **12**, 3335 (2021). doi: 10.1038/s41467-021-23614-x; pmid: 34099661
106. Z. Wu, L. Tian, X. Liu, Y. Zhang, X. Li, TIR signal promotes interactions between lipase-like proteins and ADR1-L1 receptor and ADR1-L1 oligomerization. *Plant Physiol.* **187**, 681–686 (2021). doi: 10.1093/plphys/kiab305; pmid: 34608964
107. D. Lapin *et al*., A coevolved EDS1-SAG101-NRG1 module mediates cell death signaling by TIR-domain immune receptors. *Plant Cell* **31**, 2430–2455 (2019). doi: 10.1105/tpc.19.00118; pmid: 31311833
108. Z. Duxbury *et al*., Induced proximity of a TIR signaling domain on a plant-mammalian NLR chimera activates defense in plants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18832–18839 (2020). doi: 10.1073/pnas.2001185117; pmid: 32709746
109. Z. Hu *et al*., Structural and biochemical basis for induced self-propagation of NLRC4. *Science* **350**, 399–404 (2015). doi: 10.1126/science.aac5489; pmid: 26449475
110. L. Zhang *et al*., Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404–409 (2015). doi: 10.1126/science.aac5789; pmid: 26449474
111. H. Carminatti, E. Cabib, Phosphorolysis of the pyrophosphate bond of some nucleotides. *Biochim. Biophys. Acta* **53**, 417–419 (1961). doi: 10.1016/0006-3002(61)90458-9; pmid: 13876695
112. L. Adams-Phillips, A. G. Briggs, A. F. Bent, Disruption of poly (ADP-ribosyl)ation mechanisms alters responses of *Arabidopsis* to biotic stress. *Plant Physiol.* **152**, 267–280 (2010). doi: 10.1104/pp.109.148049; pmid: 19889874
113. N. Maruta *et al*., Structural basis of NLR activation and innate immune signalling in plants. *Immunogenetics* **74**, 5–26 (2022). doi: 10.1007/s00251-021-01242-5; pmid: 34981187
114. M. D. Figley *et al*., SARM1 is a metabolic sensor activated by an increased NMN/NAD^+ ratio to trigger axon degeneration. *Neuron* **109**, 1118–1136.e11 (2021). doi: 10.1016/j.neuron.2021.02.009; pmid: 33657413
115. M. Sporny *et al*., Structural basis for SARM1 inhibition and activation under energetic stress. *eLife* **9**, e62021 (2020). doi: 10.7554/eLife.62021; pmid: 33185189
116. Y. Xu *et al*., Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* **408**, 111–115 (2000). doi: 10.1038/35040600; pmid: 11081518
117. T. Ve, S. J. Williams, B. Kobe, Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. *Apoptosis* **20**, 250–261 (2015). doi: 10.1007/s10495-014-1064-2; pmid: 25451009
118. A. M. Burroughs, D. Zhang, D. E. Schäffer, L. M. Iyer, L. Aravind, Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res.* **43**, 10633–10654 (2015). doi: 10.1093/nar/gkv1267; pmid: 26590262
119. P. Laurino *et al*., An ancient fingerprint indicates the common ancestry of Rossmann-fold enzymes utilizing different ribose-based cofactors. *PLOS Biol.* **14**, e1002396 (2016). doi: 10.1371/journal.pbio.1002396; pmid: 26938925
120. R. Martin *et al*., Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020). doi: 10.1126/science.abd9993; pmid: 33273074
121. S. Ma *et al*., Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**, eabe3069 (2020). doi: 10.1126/science.abe3069; pmid: 33273071
122. J. Wang *et al*., Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* **364**, eaav5870 (2019). doi: 10.1126/science.aav5870; pmid: 30948527
123. D. W. Summers, D. A. Gibson, A. DiAntonio, J. Milbrandt, SARM1-specific motifs in the TIR domain enable NAD^+ loss and regulate injury-induced SARM1 activation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6271–E6280 (2016). doi: 10.1073/pnas.1601506113; pmid: 27671644
124. C. Shen *et al*., Multiple domain interfaces mediate SARM1 autoinhibition. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023151118 (2021). doi: 10.1073/pnas.2023151118; pmid: 33468661
125. M. Bratkowski *et al*., Structural and mechanistic regulation of the pro-degenerative NAD hydrolase SARM1. *Cell Rep.* **32**, 107999 (2020). doi: 10.1016/j.celrep.2020.107999; pmid: 32755591

126. Y. Jiang *et al*., The NAD^+-mediated self-inhibition mechanism of pro-neurodegenerative SARM1. *Nature* **588**, 658–663 (2020). doi: 10.1038/s41586-020-2862-z; pmid: 33053563
127. M. Sporny *et al*., Structural evidence for an octameric ring arrangement of SARM1. *J. Mol. Biol.* **431**, 3591–3605 (2019). doi: 10.1016/j.jmb.2019.06.030; pmid: 31278906
128. M. Di Stefano *et al*., A rise in NAD precursor nicotinamide mononucleotide (NMN) after injury promotes axon degeneration. *Cell Death Differ.* **22**, 731–742 (2015). doi: 10.1038/cdd.2014.164; pmid: 25323584
129. G. Hogrel *et al*., Cyclic nucleotide-induced superhelical structure activates a bacterial TIR immune effector. bioRxiv 2022.05.04.490601 [Preprint] (2022); doi: 10.1101/2022.05.04.490601
130. X. Cai, H. Xu, Z. J. Chen, Prion-like polymerization in immunity and inflammation. *Cold Spring Harb. Perspect. Biol.* **9**, a023580 (2017). doi: 10.1101/cshperspect.a023580; pmid: 27881448
131. H. Xu *et al*., Structural basis for the prion-like MAVS filaments in antiviral innate immunity. *eLife* **3**, e01489 (2014). doi: 10.7554/eLife.01489; pmid: 24569476
132. T. Van Damme *et al*., Wounding stress causes rapid increase in concentration of the naturally occurring 2′,3′-isomers of cyclic guanosine- and cyclic adenosine monophosphate (cGMP and cAMP) in plant tissues. *Phytochemistry* **103**, 59–66 (2014). doi: 10.1016/j.phytochem.2014.03.013; pmid: 24735826
133. E. K. Jackson, Discovery and roles of 2′,3′-cAMP in biological systems. *Handb. Exp. Pharmacol.* **238**, 229–252 (2017). doi: 10.1007/164_2015_40; pmid: 26721674
134. H. Tian *et al*., Activation of TIR signalling boosts pattern-triggered immunity. *Nature* **598**, 500–503 (2021). doi: 10.1038/s41586-021-03987-1; pmid: 34544113
135. M. Bjornson, P. Pimprikar, T. Nürnberger, C. Zipfel, The transcriptional landscape of *Arabidopsis thaliana* pattern-triggered immunity. *Nat. Plants* **7**, 579–586 (2021). doi: 10.1038/s41477-021-00874-5; pmid: 33723429
136. B. P. M. Ngou, P. Ding, J. D. G. Jones, Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34**, 1447–1478 (2022). doi: 10.1093/plcell/koac041; pmid: 35167697
137. T. Bosanac *et al*., Pharmacological SARM1 inhibition protects axon structure and function in paclitaxel-induced peripheral neuropathy. *Brain* **144**, 3226–3238 (2021). doi: 10.1093/brain/awab184; pmid: 33964142
138. R. O. Hughes *et al*., Small molecule SARM1 inhibitors recapitulate the $SARM1^{-/-}$ phenotype and allow recovery of a metastable pool of axons fated to degenerate. *Cell Rep.* **34**, 108588 (2021). doi: 10.1016/j.celrep.2020.108588; pmid: 33406435
139. S. Geisler *et al*., Gene therapy targeting SARM1 blocks pathological axon degeneration in mice. *J. Exp. Med.* **216**, 294–303 (2019). doi: 10.1084/jem.20181040; pmid: 30642945
140. S. A. Gould *et al*., *Sarm1* haploinsufficiency or low expression levels after antisense oligonucleotides delay programmed axon degeneration. *Cell Rep.* **37**, 110108 (2021). doi: 10.1016/j.celrep.2021.110108; pmid: 34910914
141. J. Gilley *et al*., Enrichment of *SARM1* alleles encoding variants with constitutively hyperactive NADase in patients with ALS and other motor nerve disorders. *eLife* **10**, e70905 (2021). doi: 10.7554/eLife.70905; pmid: 34796871
142. A. J. Bloom *et al*., Constitutively active SARM1 variants that induce neuropathy are enriched in ALS patients. *Mol. Neurodegener.* **17**, 1 (2022). doi: 10.1186/s13024-021-00511-x; pmid: 34991663
143. E. J. Brace *et al*., Distinct developmental and degenerative functions of SARM1 require NAD^+ hydrolase activity. *PLOS Genet.* **18**, e1010246 (2022). doi: 10.1371/journal.pgen.1010246; pmid: 35737728
144. K. A. Herrmann *et al*., Divergent signaling requirements of dSARM in injury-induced degeneration and developmental glial phagocytosis. *PLOS Genet.* **18**, e1010257 (2022). doi: 10.1371/journal.pgen.1010257; pmid: 35737721
145. N. D. Peterson *et al*., Pathogen infection and cholesterol deficiency activate the *C. elegans* p38 immune pathway through a TIR-1/SARM1 phase transition. *eLife* **11**, e74206 (2022). doi: 10.7554/eLife.74206; pmid: 35098926
146. M. Lukacs *et al*., Severe biallelic loss-of-function mutations in nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2) in two fetuses with fetal akinesia deformation sequence. *Exp. Neurol.* **320**, 112961 (2019). doi: 10.1016/j.expneurol.2019.112961; pmid: 31136762
147. P. Huppke *et al*., Homozygous *NMNAT2* mutation in sisters with polyneuropathy and erythromelalgia. *Exp. Neurol.* **320**, 112958 (2019). doi: 10.1016/j.expneurol.2019.112958; pmid: 31132363

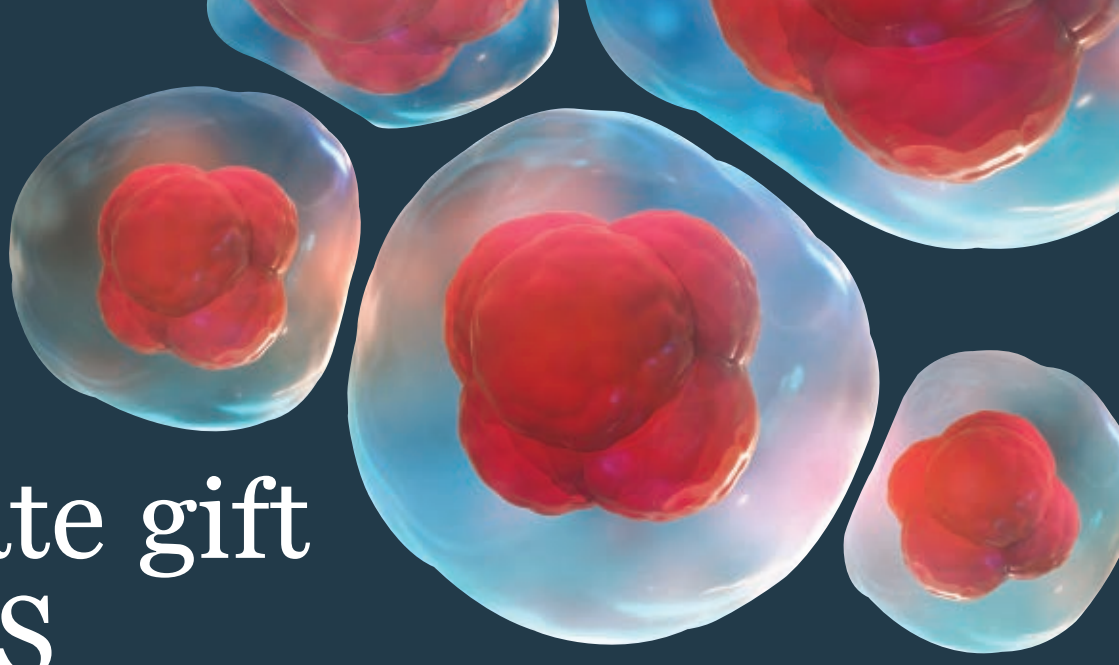
ACKNOWLEDGMENTS

We thank A. Bayless, A. DiAntonio, B. Kobe, M. Grant, Y. Sasaki, and L. Wan for editing and discussion of the manuscript. This article is subject to Howard Hughes Medical Institute (HHMI)'s Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY4.0 license immediately upon publication. Figures were created with BioRender.com and ChemDraw. **Funding:** This work was supported by the National Institutes of Health (grants RF1AG013730, R01NS087632, and R01NS119812 to J.M.), the National Science Foundation (grant IOS-1758400 to J.L.D. and M.T.N.), and HHMI (J.L.D.). J.L.D. is an HHMI investigator. M.T.N. received funding from Syngenta Crop Protection, LLC. **Competing interests:** J.M., cofounder of Disarm Therapeutics, and K.E. may derive income from licensing technology to Disarm Therapeutics, a wholly owned subsidiary of Eli Lilly and Company. J.M. and K.E. are inventors on patent WO2018057989A1 that describes SARM1 inhibitors as a treatment for neurodegenerative disease.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by HHMI, a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

Submitted 10 April 2022; accepted 24 June 2022
Published online 7 July 2022
10.1126/science.abo0001

1848 SOCIETY
AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

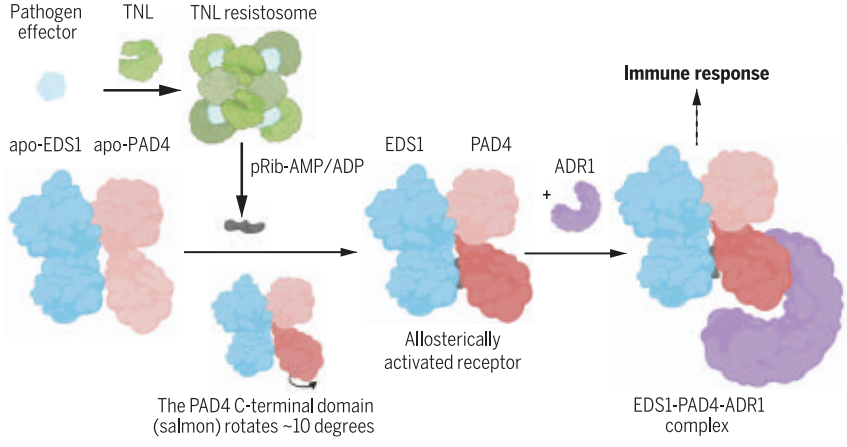
PLANT SCIENCE

Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity

Shijia Huang†, Aolin Jia†, Wen Song†, Giuliana Hessler†, Yonggang Meng†, Yue Sun, Lina Xu, Henriette Laessle, Jan Jirschitzka, Shoucai Ma, Yu Xiao, Dongli Yu, Jiao Hou, Ruiqi Liu, Huanhuan Sun, Xiaohui Liu, Zhifu Han*, Junbiao Chang*, Jane E. Parker*, Jijie Chai*

INTRODUCTION: Nucleotide-binding leucine-rich repeat (NLR) proteins are intracellular immune receptors with roles in both plant and animal protection against diseases. Plant NLRs can be divided into two broad subclasses: CNLs with an N-terminal coiled-coil domain and TNLs with an N-terminal Toll/interleukin-1 receptor (TIR) domain. Plant NLRs confer specific recognition of pathogen effectors to initiate effector-triggered immunity. NLR recognition of pathogen effectors induces NLR oligomerization, leading to large complexes called resistosomes that mediate effector-triggered immunity signaling. The TNL resistosomes, *Arabidopsis* RPP1 and *Nicotiana benthamiana* Roq1, act as nicotinamide adenine dinucleotide hydrolases (NADases) encoded in the N-terminal TIR domain. TNL-dependent immune signaling requires the conserved lipase-like proteins Enhanced Disease Susceptibility 1 (EDS1) and its two direct and exclusive partners Phytoalexin Deficient 4 (PAD4) and Senescence-Associated Gene 101 (SAG101), along with the helper NLR subfamilies Activated Disease Resistance 1 (ADR1) and N requirement gene 1 (NRG1) of the CNL class. Multiple lines of evidence support the functional cooperation of EDS1-PAD4 dimers with ADR1 and EDS1-SAG101 dimers with NRG1 in conferring immunity. An open question is how EDS1 dimers and helper NLRs integrate signals from the TNL resistosomes to initiate plant defense.

RATIONALE: TIR domain NADase-catalyzed products are hypothesized to be perceived by EDS1-PAD4 and EDS1-SAG101 dimers to activate, respectively, ADR1 and NRG1. However, the molecular identities and roles of the presumed TIR-generated signals remain elusive. Plant TIR domain proteins are known to catalyze the production of multiple nucleotide-based compounds, making it challenging to identify the precise products recognized by EDS1 heterodimers. A TNL-triggered EDS1/SAG101/NRG1 signaling module was successfully reconstituted in a heterologous *Nb* tobacco system. This suggests that TIR-catalyzed small molecules might also be synthesized in a heterologous eukaryotic cell line, such as insect cells. We hypothesized that EDS1-PAD4 and EDS1-SAG101 are receptors for specific TIR-catalyzed products. Thus, if a TNL resistosome is coexpressed with EDS1 and PAD4 or EDS1 and SAG101 in insect cells, then the TNL-generated small molecules will be likely captured by the EDS1 dimers and identified.



TIR-catalyzed signaling molecules activate plant immunity. Pathogen effector activation induces assembly of a TNL resistosome with TIR catalytic activity producing pRib-AMP and pRib-ADP. These two small molecules bind specifically to a pocket between EDS1 and PAD4, triggering a rotation of ~10° in the PAD4 C-terminal domain. This creates a fully active configuration of the ligand-binding pocket between EDS1 and PAD4 and allosterically induces EDS1-PAD4 association with a CNL-type helper NLR, ADR1, thereby potentiating an ADR1-mediated immune response.

RESULTS: Coexpression of the RPP1 resistosome (containing RPP1 with its cognate *Hyaloperonospora arabidopsidis* effector ATR1), EDS1 and PAD4 (collectively called RAEP), and ADR1-like 1 (ADR1-L1) resulted in EDS1-PAD4 interaction with ADR1-L1 in insect cells. A similar activity was also demonstrated for the monocot TIR from *Brachypodium distachyon*. The induced EDS1-PAD4 interaction with ADR1-L1 was abrogated by an NADase catalytic mutation (E158A) of RPP1. Compounds extracted from EDS1-PAD4 protein purified from the RAEP system stimulated apo-EDS1-PAD4 binding to ADR1-L1. A crystal structure of the EDS1-PAD4 complex revealed that a small molecule, 2′-(5″-phosphoribosyl)-5′-adenosine diphosphate (pRib-ADP), binds specifically to a conserved pocket between EDS1 and PAD4. Liquid chromatography–high resolution mass spectrometry validated the identity of pRib-ADP and identified 2′-(5″-phosphoribosyl)-5′-adenosine monophosphate (pRib-AMP) in the EDS1-PAD4 complex. As further confirmation of the two compounds, synthetic pRib-AMP and pRib-ADP exhibited potent activity in inducing the EDS1-PAD4 interaction with ADR1-L1, but were much less efficient in promoting the EDS1-SAG101 interaction with NRG1A. Expression of the TIR-only protein *Arabidopsis* RBA1 in *Nb* tobacco promoted the accumulation of pRib-AMP. The pRib-AMP/ADP–binding pocket overlaps with a previously identified surface groove of EDS1-PAD4 that is required for pathogen resistance. Structural comparison between pRib-ADP–bound and apo-EDS1-PAD4 complexes revealed conformational changes in the PAD4 C-terminal domain, which in turn results in a fully active configuration of the pRib-ADP–binding site between the EP domains of PAD4 and EDS1.

CONCLUSION: The study identifies pRib-AMP and pRib-ADP as TIR-catalyzed bioactive products in diverse seed plants and establishes EDS1-PAD4 as a preferential receptor complex for these two small molecules. pRib-AMP and pRib-ADP binding involves an induced-fit mechanism and allosterically induces an EDS1-PAD4 association with ADR1 to promote immunity signaling. Our data provide a missing link between pathogen perception and defense. They further suggest that pRib-AMP and pRib-ADP are a class of second messengers regulating ADR1-mediated plant immunity. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: hanzhifu@mail.tsinghua.edu.cn (Z.H.); changjunbiao@zzu.edu.cn (J.Chang); parker@mpipz.mpg.de (J.E.P.); chaijj@tsinghua.edu.cn (J.Chai)
†These authors contributed equally to this work.
Cite this article as S. Huang *et al.*, *Science* **377**, eabq3297 (2022). DOI: 10.1126/science.abq3297

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq3297

RESEARCH ARTICLE

PLANT SCIENCE

Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity

Shijia Huang[1]†, Aolin Jia[1]†, Wen Song[2,3]†, Giuliana Hessler[3]†, Yonggang Meng[4]†, Yue Sun[1], Lina Xu[5], Henriette Laessle[3], Jan Jirschitzka[2,3], Shoucai Ma[1], Yu Xiao[1], Dongli Yu[2,3], Jiao Hou[4], Ruiqi Liu[4], Huanhuan Sun[2,3], Xiaohui Liu[5], Zhifu Han[1]*, Junbiao Chang[4,6]*, Jane E. Parker[3]*, Jijie Chai[1,2,3]*

Plant nucleotide-binding leucine-rich repeat-containing (NLR) receptors with an N-terminal Toll/interleukin-1 receptor (TIR) domain sense pathogen effectors to enable TIR-encoded nicotinamide adenine dinucleotide hydrolase (NADase) activity for immune signaling. TIR-NLR signaling requires the helper NLRs N requirement gene 1 (NRG1), Activated Disease Resistance 1 (ADR1), and Enhanced Disease Susceptibility 1 (EDS1), which forms a heterodimer with each of its paralogs Phytoalexin Deficient 4 (PAD4) and Senescence-Associated Gene 101 (SAG101). Here, we show that TIR-containing proteins catalyze the production of 2′-(5″-phosphoribosyl)-5′-adenosine monophosphate (pRib-AMP) and diphosphate (pRib-ADP) in vitro and in planta. Biochemical and structural data demonstrate that EDS1-PAD4 is a receptor complex for pRib-AMP and pRib-ADP, which allosterically promote EDS1-PAD4 interaction with ADR1-L1 but not NRG1A. Our study identifies TIR-catalyzed pRib-AMP and pRib-ADP as a missing link in TIR signaling through EDS1-PAD4 and as likely second messengers for plant immunity.

Nucleotide-binding leucine-rich repeat-containing (NLR) proteins are intracellular immune receptors with crucial roles in innate immunity in plants and animals (*1*). Plant NLRs have a conserved organization with a central nucleotide-binding domain (NBD), a C-terminal leucine-rich repeat (LRR) domain, and an N-terminal coiled-coil (CC) or Toll/interleukin-1 receptor (TIR) domain, referred to as CNL or TNL receptors, respectively. Monocot plants lack TNLs but encode TIR-only and TIR-NB proteins (*2*, *3*). Plant NLRs mediate specific recognition of pathogen effectors to initiate effector-triggered immunity (ETI). Activated ETI signaling results in pathogen restriction and often a host hypersensitive cell death response at infection sites (*4*). Direct or indirect recognition of effectors induces NLR oligomerization (*5*), resulting in the formation of large protein complexes called resistosomes (*1*, *6*, *7*). The *Arabidopsis* CNL ZAR1 resistosome functions as a calcium-permeable channel for ETI signaling (*8*). By contrast, the TNL resistosomes *Arabidopsis* RPP1 (*9*) and *Nicotiana benthamiana* (*Nb* tobacco) Roq1 (*10*) are nicotinamide adenine dinucleotide hydrolase (NADase) enzymes encoded by their N-terminal TIR domain (*11*, *12*). TIR NADase activity relies on the presence of a conserved glutamate residue that is also essential for TIR-triggered immune signaling (*11*, *12*).

Plant TIR-dependent immune signaling requires downstream conserved helper NLRs (hNLRs) of the CNL class and the lipase-like protein family members Enhanced Disease Susceptibility 1 (EDS1), Phytoalexin Deficient 4 (PAD4), and Senescence-Associated Gene 101 (SAG101) (*13*, *14*). The hNLRs divide into the ADR1 [ADR1, ADR1-like 1 (ADR1-L1), and ADR1-L2)] and NRG1 (NRG1A, NRG1B, and NRG1C) subfamilies in *Arabidopsis*. EDS1 exclusive heterodimers with PAD4 and SAG101 confer basal immunity and ETI responses (*15*–*17*). The C-terminal EP domain (shared by EDS1, PAD4, and SAG101) contains positively charged residues required for TNL signaling (*16*, *18*–*20*). Phylogenomic (*16*, *21*) and genetic (*16*, *17*, *19*, *22*–*25*) studies suggest a functional partnership of EDS1-PAD4 with ADR1 and EDS1-SAG101 with NRG1. This is supported by induced associations of *Arabidopsis* EDS1-SAG101 and EDS1-PAD4 with NRG1 and ADR1, respectively (*19*, *26*) in planta, which are required for pathogen resistance and potentially Ca^{2+}-channel activity of hNLRs (*19*, *27*). In the current model, products generated by the TIR-encoded NADase activity are intercepted by EDS1 heterodimers for activation of hNLRs (*19*, *26*). Variant-cyclic ADP-ribose (v-cADPR) is a TIR-generated product (*11*, *12*, *28*), although its molecular identity and role in the plant immune system remain elusive. Thus, TIR-catalyzed chemical products mediating immune signaling are unknown.

Here, we identified plant TIR-catalyzed 2′-(5″-phosphoribosyl)-5′-adenosine monophosphate (pRib-AMP) and diphosphate (pRib-ADP) through mass spectrometry (MS) and structural biology. These two compounds are produced by expression of the RPP1 resistosome or TIR domain of *Arabidopsis* RPS4 in insect cells. They bind to and promote EDS1-PAD4 direct interaction with ADR1-L1. The same activities are reproduced with chemically synthesized pRib-AMP and pRib-ADP. These molecules are much less efficient in promoting the EDS1-SAG101 interaction with NRG1A, indicating preferential recognition of pRib-AMP and pRib-ADP by EDS1-PAD4. Structural comparisons reveal that pRib-ADP binding results in conformational changes in the EP domain of PAD4, enabling interaction with ADR1-L1. The data establish two nucleotide derivatives, pRib-AMP and pRib-ADP, as TIR-produced signals activating the EDS1/PAD4/ADR1 immunity module in plants. Identification of pRib-AMP and pRib-ADP as potential immune second messengers opens the way for designing small molecules to manipulate plant immunity.

Reconstitution of TIR-induced EDS1-PAD4 interaction with ADR1 in insect cells

A TNL-triggered *Arabidopsis* EDS1/SAG101/NRG1 immunity signaling module was reconstituted in *Nb* tobacco (*16*, *19*). To facilitate identification of TIR domain catalytic products, we tested whether NRG1 or ADR1 activation could be reconstituted in heterologous cells. We coexpressed *Arabidopsis* RPP1, its activating pathogen effector ATR1/EDS1/PAD4 (collectively called RAEP) and ADR1-L1 in insect cells and assayed for induced EDS1-PAD4 interaction with ADR1-L1 using a pull-down assay (*16*, *26*). As shown in Fig. 1A (top) and fig. S1A, EDS1-PAD4 but not EDS1-SAG101 copurified with ADR1-L1 in the assay, supporting specific and direct interaction of EDS1-PAD4 with ADR1-L1. The interaction was dependent on the presence of the RPP1 resistosome (fig. S2). EDS1-SAG101 but not EDS1-PAD4 interacted with NRG1A when coexpressed with RPP1 and ATR1 (called RAES) (Fig. 1A, bottom, and fig. S1A). These interactions were abolished when using an NADase catalytic mutant (E158A) of RPP1 (*9*) (Fig. 1A), suggesting that RPP1 TIR catalytic products promote EDS1 heterodimer associations with specific hNLR subtypes. The isolated TIR domain (residues 1 to 236) of *Arabidopsis* TNL RPS4 triggers *EDS1*-dependent HR cell death when expressed in *Nicotiana tabacum* plants (*29*). Insect cell coexpression of RPS4 TIR also promoted EDS1-PAD4 and EDS1-SAG101 binding to ADR1-L1

[1]Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, Center for Plant Biology, School of Life Sciences, Tsinghua University, 100084 Beijing, China. [2]Institute of Biochemistry, University of Cologne, 50674 Cologne, Germany. [3]Max-Planck Institute for Plant Breeding Research, Department of Plant-Microbe Interactions, 50829 Cologne, Germany. [4]School of Pharmaceutical Sciences, Zhengzhou University, 450001 Zhengzhou, China. [5]National Protein Science Facility, Tsinghua University, 100084 Beijing, China. [6]Henan Key Laboratory of Organic Functional Molecules and Drug Innovation, Henan Normal University, 453007 Xinxiang, China.
*Corresponding author. Email: hanzhifu@mail.tsinghua.edu.cn (Z.H.); changjunbiao@zzu.edu.cn (J.Chang); parker@mpipz.mpg.de (J.E.P.); chaijj@tsinghua.edu.cn (J.Chai)
†These authors contributed equally to this work.

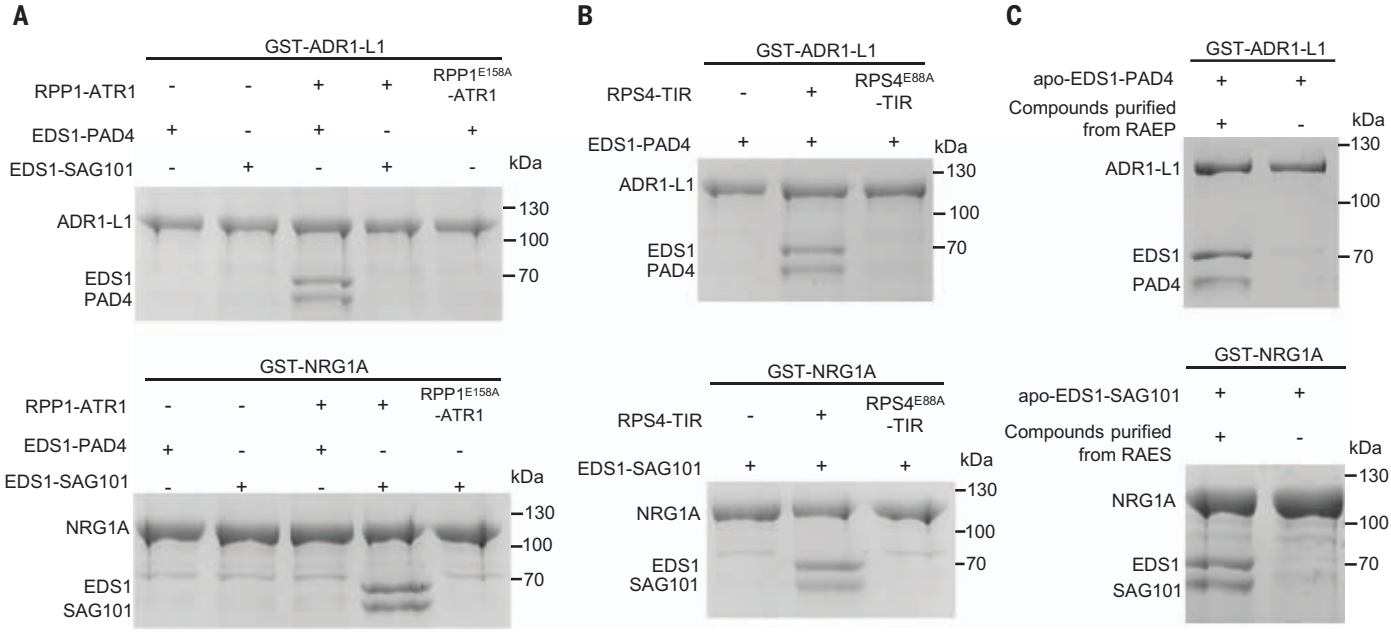


Fig. 1. TIR-catalyzed products induce the EDS1 heterodimer interaction with hNLRs in insect cells. (**A**) RPP1-ATR1 induces an EDS1-PAD4 and EDS1-SAG101 interaction with ADR1-L1 and NRG1A, respectively, in insect cells. N-terminally GST-tagged ADR1-L1 or NRG1A was coexpressed with nontagged EDS1, C-terminally strep-tagged PAD4 or SAG101, C-terminally strep-tagged RPP1 (WT or catalytic mutant E158A), and 10×His-tagged ATR1 in insect cells. GST-ADR1-L1 or NRG1A was purified using glutathione sepharose 4B (GS4B) beads. GS4B-bound proteins were eluted, separated by SDS–polyacrylamide gel electrophoresis (SDS-PAGE), and detected by Coomassie brilliant blue staining. (**B**) RPS4 TIR (residues 1 to 236) induces an EDS1-PAD4 and EDS1-SAG101 interaction with ADR1-L1 and NRG1A, respectively, in insect cells. Experiments were performed as described in (A) except that RPP1 and ATR1 were replaced with RPS4 TIR. (**C**) EDS1-PAD4 or EDS1-SAG101 was expressed as described in (A) and purified with Strep-Tactin resin. The purified protein was concentrated and denatured by heating. After centrifugation, the supernatant was collected and incubated with separately purified N-terminally GST-tagged ADR1-L1 or NRG1A and apo-EDS1-PAD4 or apo-EDS1-SAG101. The mixture was passed over GS4B beads, and bound proteins were eluted, analyzed by SDS-PAGE, and detected by Coomassie brilliant blue staining.

and NRG1A, respectively, which was abolished by mutation of the RPS4 TIR catalytic residue Glu^{88} (Fig. 1B and fig. S1B). These results show that TIR domain catalytic products induce EDS1-PAD4 and EDS1-SAG101 interactions with their cofunctioning hNLRs. A TIR-only protein from the monocot species *Brachypodium distachyon* (*Bd*TIR) also elicits EDS1-dependent cell death when expressed in *Nb* tobacco (*12*). Coexpression with *Bd*TIR in insect cells similarly induced EDS1-PAD4 association with ADR1-L1 (fig. S3). By contrast, a TIR-containing bacterial effector (HopAM1), which suppresses plant immune responses in an NADase-dependent manner (*30*), did not induce an ADR1-L1 interaction with EDS1-PAD4 dimer in this assay (fig. S3). These data suggest that TIR proteins from unrelated seed plant species can directly induce EDS1 dimer-hNLR associations.

We next investigated whether products of TIR enzyme activity bind to and support EDS1-PAD4 and EDS1-SAG101 interactions with hNLRs in vitro. For this, we coexpressed RAEP in insect cells. The expressed EDS1-PAD4 complex was predicted to contain products catalyzed by the RPP1 resistosome. We purified the EDS1-PAD4 protein and extracted the presumed small molecules by denaturing the complex protein. The extracted small molecules induced interaction of separately purified apo-EDS1-PAD4 with ADR1-L1 (Fig. 1C and fig. S1C). Similarly, small molecules extracted from the RAES-purified and denatured EDS1-SAG101 stimulated interaction of separately purified apo-EDS1-SAG101 with NRG1A (Fig. 1C and fig. S1C). These biochemical data show that enzymatic products of TIR proteins interact with and facilitate EDS1-PAD4– and EDS1-SAG101–specific interactions with hNLRs.

Structures of TIR-catalyzed small molecules activating EDS1-PAD4

We next determined the identities of the extracted small molecules promoting the EDS1-PAD4 interaction with ADR1-L1 using liquid chromatography coupled with high resolution mass spectrometry (LC-HRMS). Two singly charged ions with mass-to-charge ratios (m/z) 558.0657 ($z = 1^-$) and 638.0331 ($z = 1^-$) were observed in the compounds extracted from RAEP but not control samples (fig. S4A). To determine the chemical structures of these two compounds, we purified EDS1-PAD4 by coexpressing RAEP in insect cells, crystallized the EDS1-PAD4 complex and solved its structure. The crystal structure of the complex was solved with molecular replacement using apo-EDS1-SAG101 (*15*) as the template, and was refined to 2.29 Å resolution (Fig. 2A and table S1).

In the final electron density map, there is a segment of unambiguous electron density between the partner EP domains that remains unoccupied after modeling of EDS1 and PAD4 (Fig. 2B). The ion feature with m/z 558.0657 ($z = 1^-$) from LC-HRMS suggested adenosine diphosphate ribose (ADPR, molecular mass 559.0717, 2.4 ppm mass shift) within 3 ppm as one of the small molecules. However, ADPR did not match the unfilled electron density, suggesting that the small molecule bound by EDS1-PAD4 is an ADPR isomer or isomer derivative. This is consistent with LC-HRMS analysis showing that ADPR had a retention time different from that of the small molecule with m/z 558.0657 ($z = 1^-$) (fig. S4B). One ADPR isomer, pRib-AMP, fits the electron density well except where a small portion connecting the terminal phosphate group remains void (fig. S5A). The ion feature with m/z 638.0331 ($z = 1^-$), with a mass shift of 79.967 from m/z 558.0657 ($z = 1^-$), suggested a molecule with an additional phosphate group on pRib-AMP. Adding one phosphate group to the terminal void density produced pRib-ADP, which fits into the electron density map well (Fig. 2B).

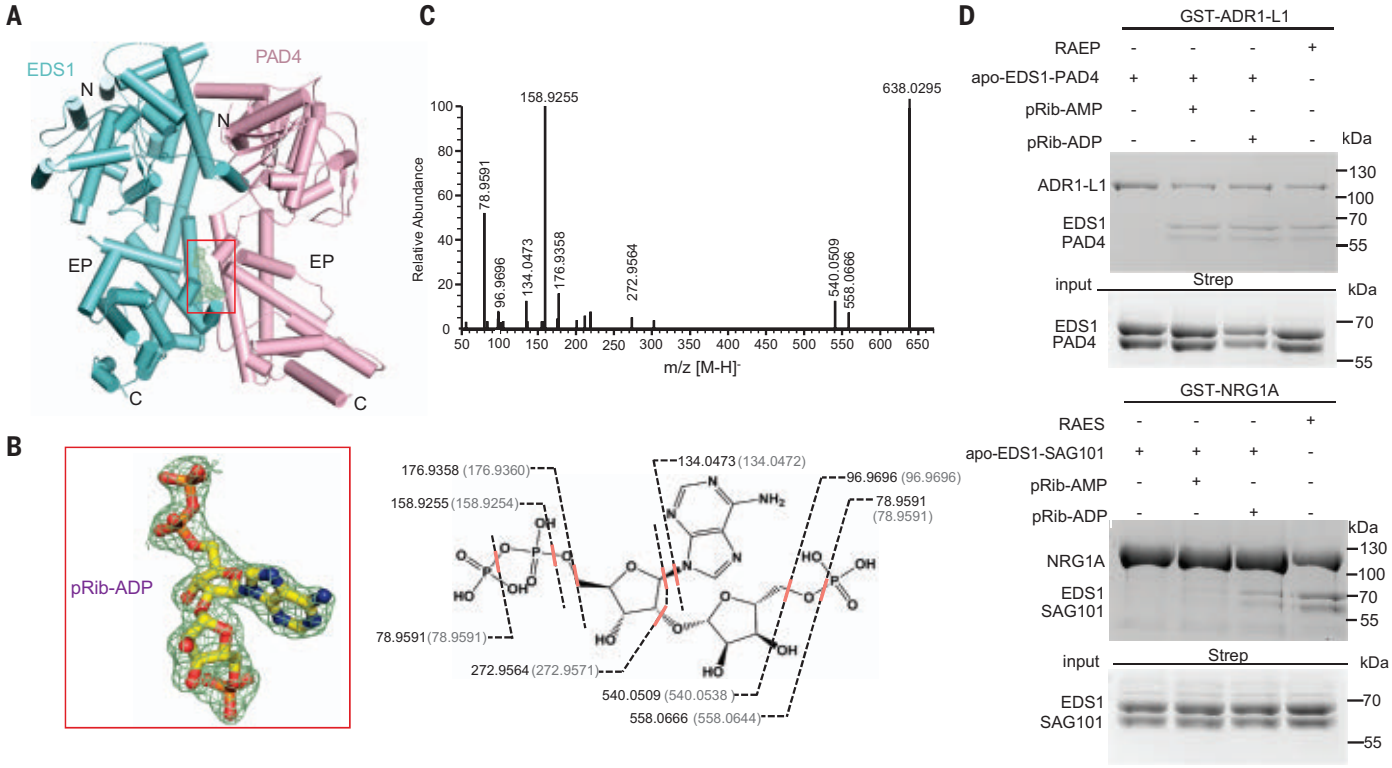


Fig. 2. Structural determination of the EDS1-PAD4 complex with bound pRib-ADP. (**A**) Crystal structure of EDS1-PAD4 bound by an RPP1-catalyzed small molecule. Highlighted within the red frame is omit electron density (*2Fo-Fc* contoured at 1.5 sigma) between the EP domains of EDS1 and PAD4. (**B**) Close-up view of the electron density highlighted in (A) with modeled pRib-ADP (in stick). (**C**) Mass analysis of small molecules bound to EDS1-PAD4. Top: MS/MS spectra after CID fragmentation of the ion with m/z = 638.0295 ($z = 1^-$). Bottom: Scheme showing the proposed routes of formation of various ions shown above. Numbers in parentheses indicate theoretical molecular weights of proposed ions. (**D**) Synthetic pRib-AMP and pRib-ADP induce a strong EDS1-PAD4 interaction and a weak EDS1-SAG101 interaction with ADR1-L1 and NRG1A, respectively. GS4B beads bound by GST-ADR1-L1 or GST-NRG1A was incubated with excess purified apo-EDS1-PAD4 or apo-EDS1-SAG101 and synthetic pRib-AMP and pRib-ADP on ice for 30 min. After washing, proteins bound to GS4B beads were eluted, analyzed by SDS-PAGE, and detected by Coomassie brilliant blue staining.

A higher pRib-ADP affinity for EDS1-PAD4 might explain why pRib-ADP was captured in the crystal structure of EDS1-PAD4.

To verify the assignment of pRib-AMP and pRib-ADP in the EDS1-PAD4 complex, standard compounds were chemically synthesized (see data S1). MS/MS spectra and chromatographic elution profiles confirmed the identification of pRib-AMP and pRib-ADP as EDS1-PAD4 ligands (Fig. 2C and fig. S5, B to D). We tested whether synthetic pRib-AMP and pRib-ADP induce EDS1-PAD4 and EDS1-SAG101 interactions, respectively, with ADR1-L1 and NRG1A. Both compounds promoted EDS1-PAD4 interaction with ADR1-L1, similar to that detected after coexpression of ADR1-L1 with RAEP (Fig. 2D, top). The small molecules were much less efficient in inducing the EDS1-SAG101 interaction with NRG1A (Fig. 2D, bottom). The synthetic pRib-AMP also promoted the EDS1-PAD4 interaction with ADR1-L1 expressed in *Nb* tobacco after pRib-AMP addition to leaf total extracts (fig. S5E). These structural and biochemical data demonstrate that pRib-AMP and pRib-ADP are TIR-catalyzed metabolites that preferentially induce the EDS1-PAD4 interaction with ADR1-L1.

Recognition mechanism of pRib-ADP by EDS1-PAD4

In the crystal structure, pRib-ADP adopts an extended conformation interacting with a pocket formed by the EDS1 and PAD4 EP domains (Fig. 3A). In support of the structural observation, the pRib-ADP binding pocket partially overlaps with positively charged EP-domain surfaces of EDS1 and PAD4 that are indispensable for EDS1 immunity signaling (*15*, *16*, *20*). pRib-ADP binding results in complete burial of the small molecule except for its terminal phosphate group (Fig. 3A). An extensive network of hydrogen bonds, including those mediated by several well-ordered water molecules, appears to govern binding of pRib-ADP to the pocket (Fig. 3B), indicating specific recognition of pRib-ADP by EDS1-PAD4. The 5-phospho-beta-D-ribosyl moiety of pRib-ADP is recognized by EDS1 residues Asp^{469}, Asn^{472}, and Tyr^{473}. The phosphate group of this moiety also forms salt bridges with Arg^{314} and Arg^{425} and van der Waals contact with Phe^{387} of PAD4 (Fig. 3B). Thus, both EDS1 and PAD4 make substantial contributions to recognition of the ADP moiety of pRib-ADP. The adenine group stacks against PAD4 Tyr^{383} at one side and EDS1 Tyr^{473} at the other, and is further stabilized by hydrogen bonds with EDS1 Thr^{482}. Besides making van der Waals contacts with PAD4 Lys^{379}/Lys^{380}, the ADP moiety ribose group hydrogen bonds with EDS1 Arg^{493} directly and through a water molecule. The two phosphate groups of the ADP moiety form in total eight hydrogen bonds with their neighboring residues - five with EDS1 and three with PAD4 (Fig. 3B, right). Structure-based sequence alignments indicated that EDS1 and PAD4 residues involved in recognition of pRib-ADP are highly conserved in different plant species (fig. S6), suggesting that pRib-AMP and pRib-ADP binding is an evolutionarily conserved activity of EDS1-PAD4. A cyclic ADPR is less likely to fit in the pRib-ADP-binding pocket in EDS1-PAD4, although the precise chemical structure of v-cADPR (*12*) remains undefined. In

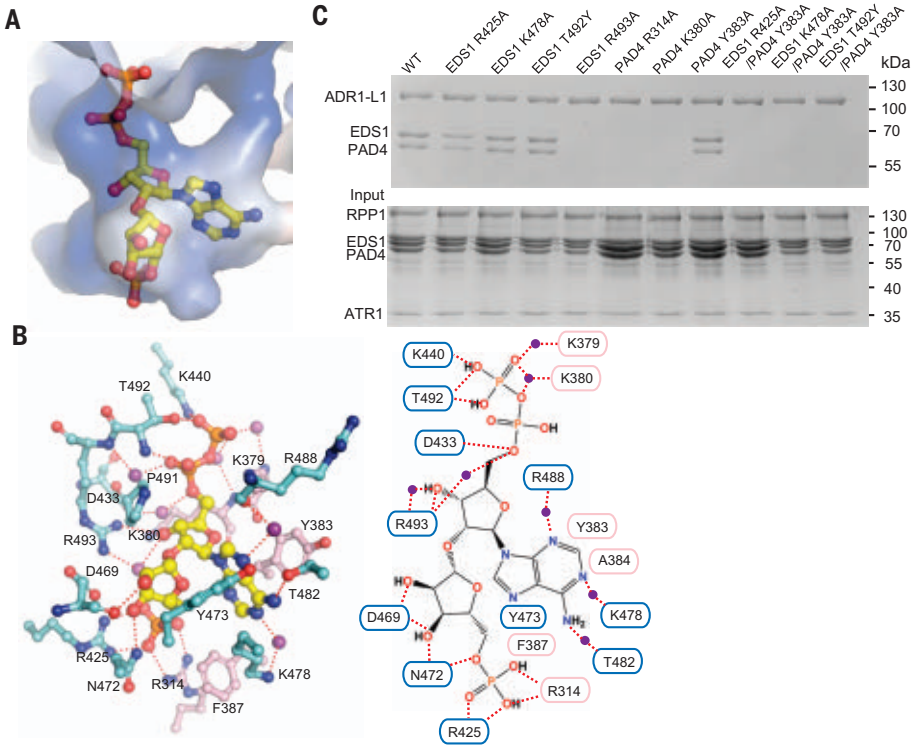


Fig. 3. Specific recognition of pRib-ADP by EDS1-PAD4. (**A**) Close-up view of the pRib-ADP–binding pocket of EDS1-PAD4. EDS1-PAD4 is shown in electrostatic surface and pRib-ADP in stick. (**B**) Left: EDS1-PAD4 EP-domain amino acid coordinates for pRib-ADP. Dashed lines represent polar interactions. Purple spheres represent water molecules. Right: Diagram showing the interaction between EDS1-PAD4 and pRib-ADP. Cyan and purple frames represent residues from EDS1 and PAD4, respectively. (**C**) Mutagenesis analysis of the EDS1-PAD4 pRib-ADP binding pocket. Assays were performed as described in Fig. 1A.

Arabidopsis, mutations in EP-domain surface residues EDS1 Arg^{493}, Lys^{478} and PAD4 Arg^{314}, Lys^{380} identified here as contacts for pRib-ADP binding, did not disrupt heterodimer formation but abolished TNL effector-triggered and basal immunity (*16*, *18*, *20*). Also, mutation of PAD4 Arg^{314} disabled TNL-induced EDS1-PAD4 association with ADR1 in *Arabidopsis* (*20*).

To verify the pRib-ADP interaction with EDS1-PAD4 observed in the crystal structure, we engineered amino acid substitutions at the pRib-ADP binding pocket between EDS1 and PAD4 and tested their effects on the EDS1-PAD4 interaction with ADR1 induced by the RPP1 resistosome in insect cells. Mutations of positively charged residues EDS1 Arg^{493}, PAD4 Arg^{314} and PAD4 Lys^{380} coordinating the phosphate group of the pRib moiety had little effect on EDS1-PAD4 dimer formation but markedly reduced the EDS1-PAD4 interaction with ADR1-L1 (Fig. 3C). When combined with EDS1 R425A, EDS1 K478A, or EDS1 T492Y, mutation of PAD4 Tyr^{383} stacking the adenosine moiety resulted in a complete loss of the interaction.

pRib-AMP and pRib-ADP accumulation in plants

We investigated whether activation of TIR-containing proteins leads to pRib-AMP and pRib-ADP production in plants. *Arabidopsis* TIR-only protein RBA1 (*31*) was used because it promotes the EDS1-PAD4 interaction with ADR1-L1 in an *Nb* tobacco transient coexpression assay (fig. S7) (*26*). LC/MS detection of these two compounds upon RBA1 expression in a signaling-defective *Nb* tobacco *epss* line lacking all EDS1-family proteins (*Nb epss*) (*16*) did not produce clear signals, possibly due to pRib-AMP and pRib-ADP instability in leaves. Indeed, LC/MS assays showed that ~50% of pRib-AMP and >90% of pRib-ADP was degraded within 1 hour of mixing with *Nb epss* tissue lysate (fig. S8). Because binding to EDS1-PAD4 might protect pRib-AMP and pRib-ADP from hydrolysis, *Nb epss* leaves expressing RBA1 were co-sonicated with insect cells expressing EDS1 and PAD4 to detect EDS1-PAD4-captured pRib-AMP and pRib-ADP with LC/MS. Plants expressing wild-type (WT) RBA1 produced strong pRib-AMP signals in the assay (Fig. 4 and fig. S9). By contrast, expression of an RBA1 NADase catalytic mutant (E86A) substantially reduced pRib-AMP accumulation (Fig. 4 and fig. S9). We failed to observe accumulation of pRib-ADP under the same conditions. This unstable small molecule might need to bind to EDS1-PAD4 immediately upon generation to evade degradation in plants. These data show that pRib-AMP produced in TIR-activated plant tissues binds to the EDS1-PAD4 complex.

Allosteric activation of EDS1-PAD4 by pRib-AMP and pRib-ADP

Having established that pRib-AMP and pRib-ADP are ligands of EDS1-PAD4, we examined the receptor complex activation mechanism by these two compounds. For this, we purified the apo-EDS1-PAD4 heterodimer from insect cells and solved a cryo–electron microscopy (cryo-EM) structure of the complex at 2.93-Å resolution (Fig. 5A, fig. S10, and table S2). The EDS1 structure is nearly identical to that in the crystal structure of pRib-ADP–bound EDS1-PAD4 (fig. S11A). As predicted, the structures of the PAD4 N-terminal lipase-like and C-terminal EP domains largely resemble their SAG101 counterparts (fig. S11B), although relative positioning of the two structural domains is substantially different (fig. S11C). Structure alignment between the free and pRib-ADP–bound forms of EDS1-PAD4 reveals that PAD4 undergoes a marked conformation change in its C-terminal EP domain upon pRib-ADP binding (Fig. 5, B and C). Compared with the PAD4 apo form, the PAD4 EP domain in its ligand-bound state is rotated ~10 degrees counterclockwise (viewed from the bottom) (Fig. 5C). This structural change leads to dissolving of the central part of a long helix connecting the PAD4 lipase–like and EP domains, which in turn allows movement of the C-terminal half of the helix toward EDS1 (Fig. 5, A and B). Similar conformation changes occur at the PAD4 short α-helix (α15) and N-terminal portion of α17. Collectively, these changes result in the formation of the pRib-ADP–binding pocket between the EP domains of PAD4 and EDS1 (Fig. 5B), and suggest that an induced-fit mechanism underlies small-molecule binding. Except for the terminal phosphate group, pRib-ADP becomes solvent inaccessible after binding to EDS1-PAD4 (Fig. 3A). Therefore, the small molecule is unlikely to be directly involved in the interaction with ADR1-L1 but rather mediates conformational changes in the PAD4 EP domain, enabling ADR1 binding to EDS1-PAD4. We conclude that pRib-AMP or pRib-ADP binding allosterically activates the EDS1-PAD4 receptor complex to interact with ADR1-family hNLRs. On the basis of our data, we propose a model in which TIR-generated pRib-AMP and pRib-ADP induce activation of EDS1-PAD4 to allosterically promote its association with ADR1 and plant immunity (fig. S12A).

Discussion

Plant TIR NADase-generated small molecule signaling intermediates proposed to activate immune responses have been elusive. One challenge has been to develop an in vitro

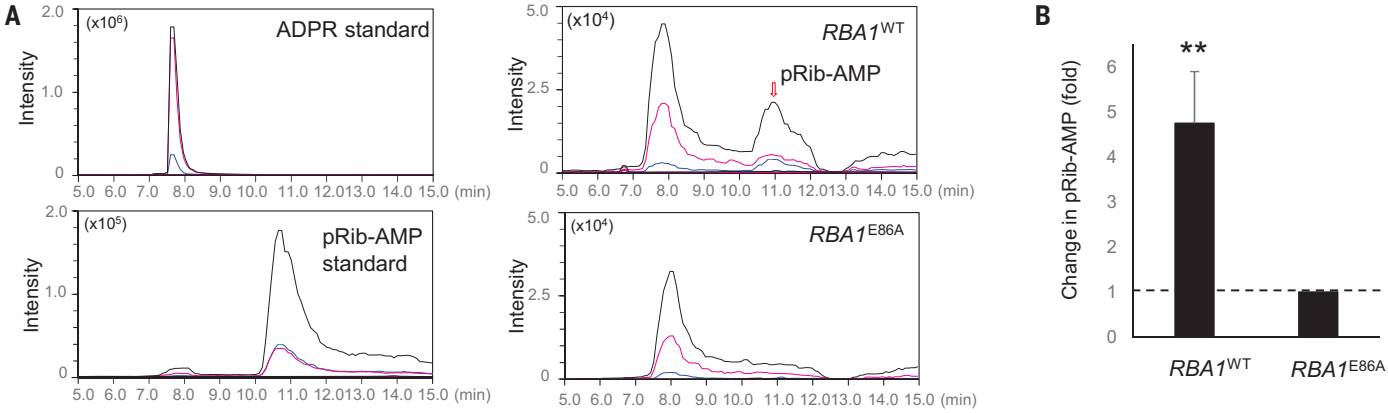


Fig. 4. Expression of TIR-only protein RBA1 induces pRib-AMP accumulation in *Nb* tobacco. (**A**) MRM analyses of pRib-AMP in *Nb* tobacco. *Nb epss* leaves expressing *RBA1* or catalytic mutant *RBA1*^{E86A} and insect cells expressing apo-EDS1-PAD4 complex were mixed and co-sonicated. The EDS1-PAD4 complex in the lysate was purified through affinity chromatography, and LC/MS was used to analyze small molecules bound in the purified EDS1-PAD4 complex. The left two panels are MRM chromatograms for ADPR and synthetic pRib-AMP, respectively. The right two panels are representative chromatograms of pRib-AMP captured by EDS1-PAD4 in *Nb epss* expressing WT *RBA1* and *RBA1*^{E86A}, respectively. Black, purple, and blue lines indicate the transitions 560.10>136.20, 560.10>348.00, and 560.10>97.20, respectively. (**B**) Relative change in pRib-AMP accumulation in *Nb epss* plants expressing *RBA1* or *RBA1*^{E86A}. The statistics were based on LC-MS results shown in (A) and fig. S9. The amount of pRib-AMP in *RBA1*^{E86A} -expressing *Nb epss* was normalized to 1.0. Significance of the difference was determined using Student's *t* test (n = 3, $^{**}P < 0.01$).

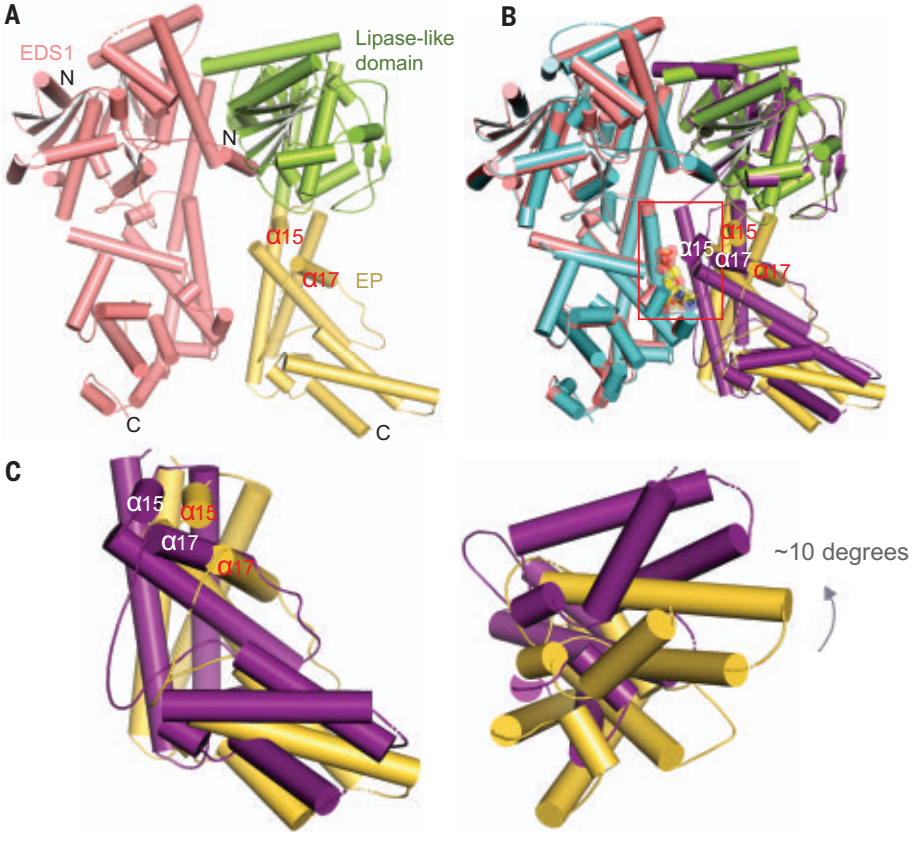


Fig. 5. Allosteric activation of EDS1-PAD4 by pRib-AMP and pRib-ADP. (**A**) Cryo-EM structure of apo-EDS1-PAD4. PAD4 lipase–like and EP domains are shown in green and light orange, respectively. (**B**) Structural alignment between apo- and pRib-ADP-bound forms of EDS1-PAD4. Highlighted with an open red frame is the pRib-ADP–binding pocket between the the EDS1 and PAD4 EP domains. EDS1 and PAD4 in the ligand-bound form are shown in cyan and purple, respectively. EDS1 in apo-EDS1-PAD4 was used as the template for structural alignment. Color codes for the apo-EDS1-PAD4 are the same as those in (A). The secondary structure elements labeled in white and red are from ligand-bound and apo-EDS1-PAD4, respectively. (**C**) Two orientations showing the structural alignment of the EP domains of PAD4 in apo- and ligand-bound forms of EDS1-PAD4. Left: the orientation is the same as that shown in (B). Right: bottom view of the structural alignment. Color codes for the apo- and pRib-ADP-bound forms of EDS1-PAD4 are the same as those in (B). The arrow shows that the PAD4 EP domain in its ligand-bound state rotates ~10° counterclockwise compared with the apo-PAD4 EP domain.

biochemical assay that detects TIR-catalyzed bioactive molecules. Here, we reconstituted EDS1 heterodimer associations with hNLRs induced by different TIR-containing proteins in insect cells (Fig. 1, A and B, and fig. S3). Expression of *Arabidopsis* TIR-only protein RBA1 in *Nb* tobacco promoted the accumulation of pRib-AMP in a catalytic activity-dependent manner (Fig. 4, A and B). These results identify pRib-AMP and pRib-ADP as potentially conserved small molecules for EDS1 signaling in seed plants. This is supported by the high degree of conservation of pRib-ADP–coordinating residues in EDS1 and PAD4 (fig. S6). Triggering of *EDS1*-dependent cell death in *Nb* tobacco plants by a monocot TIR-only protein BdTIR (*12*), along with *Bd*TIR activity promoting EDS1-PAD4-ADR1

interactions in insect cells (fig. S3), strengthen the notion that TIRs might produce pRib-AMP and pRib-ADP to activate the EDS1-PAD4-ADR1 immune signaling branch in diverse seed plant species.

The terminal phosphate group of pRib-ADP is solvent exposed when bound to EDS1-PAD4 (Fig. 3A). It is therefore possible that pRib-ADP derivatives with covalent modifications to this phosphate group are also recognized by EDS1-PAD4. Biochemical and structural data reveal that *Arabidopsis* EDS1-PAD4 functions as a receptor complex for pRib-AMP and pRib-ADP (Figs. 1 and 3). Supporting the biological significance of pRib-AMP and pRib-ADP recognition by EDS1-PAD4, mutations disrupting pRib-AMP and pRib-ADP binding (Fig. 3C) suppress EDS1-PAD4–mediated pathogen resistance (*15*, *18*, *20*). pRib-AMP and pRib-ADP binding induces conformational changes in EDS1-PAD4 to allosterically promote its association with ADR1-L1 (Fig. 5). The above characteristics lead us to propose that TIR-catalyzed products pRib-AMP and pRib-ADP act as second messengers in plant immunity signaling. The mode of action of these two compounds is reminiscent of cyclic nucleotide–based second messengers such as cyclic GMP-AMP (cGAMP) in animals (*32*) (fig. S12B).

In addition to mediating TNL receptor signaling, the *Arabidopsis* EDS1-PAD4 dimer promotes defense reprogramming downstream of a non–TNL-type of intracellular NLR RPS2 (*18*, *33*) and cell-surface immune receptor RLP23 (*14*, *34*, *35*), through what we identify here as the pRib-AMP and pRib-ADP binding pocket. Thus, pRib-AMP and pRib-ADP likely have a broad role in biotic stress signaling, consistent with the molecules being second messengers (*12*, *35*). In further comparison to well-established second messengers such as phosphatidylinositol 3,4,5-trisphosphateinositol (PIP_3) and cADPR (*36*), pRib-AMP and pRib-ADP convert externally generated stimuli (pathogen effectors) to intracellular Ca^{2+} signals, potentially by activating the Ca^{2+}-permeable channel of ADR1 (*27*). Second messenger signaling is generally terminated by metabolic enzymes (*37*). Although the processes that remove pRib-AMP and pRib-ADP are not known, these molecules are rapidly degraded in *Nb* tobacco extracts (fig. S8).

The current data show that TIR-containing proteins catalyze the production of pRib-AMP and pRib-ADP, which directly activate the EDS1-PAD4 dimer. TIR-containing proteins can also encode a 2′,3′-cAMP/cGMP synthetase activity promoting *EDS1*-dependent cell death (*38*). Thus, plant TIR-containing proteins appear to constitute a large family of enzymes producing noncanonical and nucleotide-based signaling molecules for regulation of immune and possibly other stress responses. We previously proposed that 2′,3′-cAMP/cGMP facilitate self-amplification of EDS1 signaling (*38*). An outstanding question is how TIR-generated pRib-AMP and pRib-ADP or their derivatives intersect with 2′,3′-cAMP/cGMP in driving the plant immune response.

Materials and methods summary

RPP1_WsB with a C-terminal twin-StrepII tag, ATR1_Emoy2 with a C-terminal 10×His tag, nontagged EDS1, PAD4 with a C-terminal twin-StrepII tag (collectively called RAEP), and ADR1-L1 with an N-terminal GST tag were coexpressed in Sf21 insect cells. The coexpressed proteins were purified using glutathione sepharose 4B beads, and proteins bound to the beads were eluted with glutathione. Similar methods were used to reconstitute the EDS1-SAG101 interaction with NRG1A.

The apo-EDS1-PAD4 complex protein was expressed using the above constructs in insect cells, purified using streptactin beads, and further cleaned by size exclusion chromatography. The purified apo-EDS1-PAD4 protein was used for cryo-EM analysis. EDS1-PAD4 purified by coexpression of RAEP was used for crystallization, which was performed using hanging-drop vapor diffusion. All diffraction datasets of crystals of the protein complex were collected at Shanghai Synchrotron Radiation Facility (SSRF) and processed using the HKL2000 software package. The crystal structure of the EDS1-PAD4 complex was determined by molecular replacement refined by Phenix.

Cryo-EM datasets of the apo-EDS1-PAD4 complex were collected in counted superresolution mode on a Titan Krios3 (FEI) electron microscope. The contrast transfer function was corrected using CTFFIND4. Further processing was performed using RELION-3.1. The final resolution was defined on the basis of gold standard FSC 0.143 criteria. Local resolution distribution was evaluated using RELION. Models were refined using Phenix, and model validation was performed with MolProbity and EMRinger.

Small molecules bound in EDS1-PAD4 were extracted and analyzed by LC-HRMS. Data dependent acquisition was applied for MS/MS. Data analysis was performed using Xcalibur version 4.4 software (Thermo Fisher Scientific). The LC-HRMS assays of the extracted molecules and chemically synthesized pRib-AMP and pRib-ADP were performed to determine the identities of the small molecules.

Coexpression of RAEP with WT or mutant forms of EDS1 and PAD4 was performed in insect cells, and pull-down assays were used to assess the impact of mutations on the EDS1-PAD4 interaction with ADR1-L1.

To determine the production of pRib-AMP and pRib-ADP in plants, *RBA1* WT and an *RBA1* catalytic mutant E86A with a C-terminal Myc tag were transformed into *Agrobacterium* and transiently expressed in an *Nicotiana benthamiana* CRISPR quadruple *eds1apad4 sag101asag101b* (*Nb epss*) mutant. Leaf powders of *RBA1* WT or *RBA1* E86A were harvested and mixed with insect cells expressing EDS1 and PAD4. Then EDS1-PAD4 protein was purified using Ni-NTA and the small molecules bound in the complex protein were extracted for LC-MS analysis.

REFERENCES AND NOTES

1. N. Maruta *et al*., Structural basis of NLR activation and innate immune signalling in plants. *Immunogenetics* **74**, 5–26 (2022). doi: 10.1007/s00251-021-01242-5; pmid: 34981187
2. B. C. Meyers, M. Morgante, R. W. Michelmore, TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J.* **32**, 77–92 (2002). doi: 10.1046/j.1365-313X.2002.01404.x; pmid: 12366802
3. D. Lapin, O. Johanndrees, Z. Wu, X. Li, J. E. Parker, Molecular innovations in plant TIR-based immunity signaling. *Plant Cell* **34**, 1479–1496 (2022). doi: 10.1093/plcell/koac035; pmid: 35143666
4. J. D. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016). doi: 10.1126/science.aaf6395; pmid: 27934708
5. B. P. M. Ngou, P. Ding, J. D. Jones, Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34**, 1447–1478 (2022). doi: 10.1093/plcell/koac041; pmid: 35167697
6. J. Wang, J. Chai, Molecular actions of NLR immune receptors in plants and animals. *Sci. China Life Sci.* **63**, 1303–1316 (2020). doi: 10.1007/s11427-019-1687-6; pmid: 32613490
7. Y. Xiong, Z. Han, J. Chai, Resistosome and inflammasome: Platforms mediating innate immunity. *Curr. Opin. Plant Biol.* **56**, 47–55 (2020). doi: 10.1016/j.pbi.2020.03.010; pmid: 32554225
8. G. Bi *et al*., The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell* **184**, 3528–3541. e12 (2021). doi: 10.1016/j.cell.2021.05.003; pmid: 33984278
9. S. Ma *et al*., Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**, eabe3069 (2020). doi: 10.1126/science.abe3069; pmid: 33273071
10. R. Martin *et al*., Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020). doi: 10.1126/science.abd9993; pmid: 33273074
11. S. Horsefield *et al*., NAD^+ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019). doi: 10.1126/science.aax1911; pmid: 31439792
12. L. Wan *et al*., TIR domains of plant immune receptors are NAD^+-cleaving enzymes that promote cell death. *Science* **365**, 799–803 (2019). doi: 10.1126/science.aax1771; pmid: 31439793
13. J. M. Feehan, B. Castel, A. R. Bentham, J. D. Jones, Plant NLRs get by with a little help from their friends. *Curr. Opin. Plant Biol.* **56**, 99–108 (2020). doi: 10.1016/j.pbi.2020.04.006; pmid: 32554226
14. J. A. Dongus, J. E. Parker, EDS1 signalling: At the nexus of intracellular and surface receptor immunity. *Curr. Opin. Plant Biol.* **62**, 102039 (2021). doi: 10.1016/j.pbi.2021.102039; pmid: 33930849
15. S. Wagner *et al*., Structural basis for signaling by exclusive EDS1 heteromeric complexes with SAG101 or PAD4 in plant innate immunity. *Cell Host Microbe* **14**, 619–630 (2013). doi: 10.1016/j.chom.2013.11.006; pmid: 24331460
16. D. Lapin *et al*., A Coevolved EDS1-SAG101-NRG1 Module Mediates Cell Death Signaling by TIR-Domain Immune Receptors. *Plant Cell* **31**, 2430–2455 (2019). doi: 10.1105/tpc.19.00118; pmid: 31311833
17. J. Gantner, J. Ordon, C. Kretschmer, R. Guerois, J. Stuttmann, An EDS1-SAG101 Complex Is Essential for TNL-Mediated Immunity in *Nicotiana benthamiana*. *Plant Cell* **31**, 2456–2474 (2019). doi: 10.1105/tpc.19.00099; pmid: 31266900
18. D. D. Bhandari *et al*., An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in *Arabidopsis*. *Nat. Commun.* **10**, 772 (2019). doi: 10.1038/s41467-019-08783-0; pmid: 30770836
19. X. Sun *et al*., Pathogen effector recognition-dependent association of NRG1 with EDS1 and SAG101 in TNL receptor

immunity. *Nat. Commun.* **12**, 3335 (2021). doi: 10.1038/s41467-021-23614-x; pmid: 34099661
20. J. A. Dongus *et al*., Cavity surface residues of PAD4 and SAG101 contribute to EDS1 dimer signaling specificity in plant immunity. *Plant J.* **110**, 1415–1432 (2022). doi: 10.1111/tpj.15747; pmid: 35324052
21. S. M. Collier, L. P. Hamel, P. Moffett, Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact.* **24**, 918–931 (2011). doi: 10.1094/MPMI-03-11-0050; pmid: 21501087
22. T. Qi *et al*., NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10979–E10987 (2018). doi: 10.1073/pnas.1814856115; pmid: 30373842
23. B. Castel *et al*., Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* **222**, 966–980 (2019). doi: 10.1111/nph.15659; pmid: 30582759
24. Z. Wu *et al*., Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytol.* **222**, 938–953 (2019). doi: 10.1111/nph.15665; pmid: 30585636
25. S. C. Saile *et al*., Two unequally redundant "helper" immune receptor families mediate *Arabidopsis* thaliana intracellular "sensor" immune receptor functions. *PLOS Biol.* **18**, e3000783 (2020). doi: 10.1371/journal.pbio.3000783; pmid: 32925907
26. Z. Wu, L. Tian, X. Liu, Y. Zhang, X. Li, TIR signal promotes interactions between lipase-like proteins and ADR1-L1 receptor and ADR1-L1 oligomerization. *Plant Physiol.* **187**, 681–686 (2021). doi: 10.1093/plphys/kiab305; pmid: 34608964
27. P. Jacob *et al*., Plant "helper" immune receptors are Ca^{2+}-permeable nonselective cation channels. *Science* **373**, 420–425 (2021). doi: 10.1126/science.abg7917; pmid: 34140391
28. Z. Duxbury *et al*., Induced proximity of a TIR signaling domain on a plant-mammalian NLR chimera activates defense in plants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18832–18839 (2020). doi: 10.1073/pnas.2001185117; pmid: 32709746
29. M. R. Swiderski, D. Birker, J. D. Jones, The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. *Mol. Plant Microbe Interact.* **22**, 157–165 (2009). doi: 10.1094/MPMI-22-2-0157; pmid: 19132868
30. S. Eastman *et al*., A phytobacterial TIR domain effector manipulates NAD^+ to promote virulence. *New Phytol.* **233**, 890–904 (2022). doi: 10.1111/nph.17805; pmid: 34657283
31. M. T. Nishimura *et al*., TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2053–E2062 (2017). doi: 10.1073/pnas.1620973114; pmid: 28137883
32. X. Cai, Y. H. Chiu, Z. J. Chen, The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol. Cell* **54**, 289–296 (2014). doi: 10.1016/j.molcel.2014.03.040; pmid: 24766893
33. H. Cui *et al*., Antagonism of transcription factor MYC2 by EDS1/PAD4 complexes bolsters salicylic acid defense in *Arabidopsis* effector-triggered immunity. *Mol. Plant* **11**, 1053–1066 (2018). doi: 10.1016/j.molp.2018.05.007; pmid: 29842929
34. R. N. Pruitt *et al*., The EDS1-PAD4-ADR1 node mediates *Arabidopsis* pattern-triggered immunity. *Nature* **598**, 495–499 (2021). doi: 10.1038/s41586-021-03829-0; pmid: 34497423
35. H. Tian *et al*., Activation of TIR signalling boosts pattern-triggered immunity. *Nature* **598**, 500–503 (2021). doi: 10.1038/s41586-021-03987-1; pmid: 34544113
36. J. V. Gerasimenko, M. Sherwood, A. V. Tepikin, O. H. Petersen, O. V. Gerasimenko, NAADP, cADPR and IP3 all release $Ca2^+$ from the endoplasmic reticulum and an acidic store in the secretory granule area. *J. Cell Sci.* **119**, 226–238 (2006). doi: 10.1242/jcs.02721; pmid: 16410548
37. M. Conti, J. Beavo, Biochemistry and physiology of cyclic nucleotide phosphodiesterases: Essential components in cyclic nucleotide signaling. *Annu. Rev. Biochem.* **76**, 481–511 (2007). pmid: 17376027
38. D. Yu *et al*., TIR domains of plant immune receptors are 2′,3′-cAMP/cGMP synthetases mediating cell death. *Cell* **185**, 2370–2386.e18 (2022). doi: 10.1016/j.cell.2022.04.032; pmid: 35597242

ACKNOWLEDGMENTS

We thank the National Protein Science Facility at Tsinghua University for technical assistance; the staff of Beamline BL02U1, BL19U at SSRF for assistance in data collection; J. Lei, X. Li, X. Fan, and N. Liu at Tsinghua University for data collection; the Tsinghua University Branch of the China National Center for Protein Sciences (Beijing) for providing the cryo-EM facility support and the computational facility support on the cluster of Bio-Computing Platform; and O. Johanndrees (MPI-Cologne) for providing sequences of EDS1 and PAD4 orthologs. **Funding:** This work was supported by the National Key Research and Development Program of China (grant 2021YFA1300701 to Z.H.), the Alexander von Humboldt Foundation (professorship to J.C.), the Max-Planck-Gesellschaft (J.E.P. and a Max Planck fellowship to J.C.), Deutsche Forschungsgemeinschaft (grant SFB-1403-414786233 to J.C. and J.E.P.), Germany's Excellence Strategy CEPLAS (EXC-2048/1, project 390686111 to J.C. and J.E.P.), Max-Planck International PhD Research School (IMPRS) support (G.H. and H.L.), and the National Natural Science Foundation of China (grants 82130103 and U1804283 to J.B.C.). **Author contributions:** J.C., J.E.P., J.B.C., Z.H., S.H., and A.J. designed the experiments. S.H., A.J., Y.S., W.S., Y.X., and J.C. performed the recombinant protein expression assays, purification, structure determinations, modeling, and data analysis. S.H., A.J., W.S., G.H., H.L., J.J., D.Y., L.X., X.L., S.M., H.S., Z.H., J.E.P., and J.C. performed the biochemical assays. G.H., H.L. and J.E.P. performed plant expression and Western blot analysis. Y.M., R.L., J.H., and J.B.C. synthesized the small molecules. All authors analyzed the data. J.C. wrote the manuscript and J.E.P., J.B.C., Z.H., S.H., and A.J. edited the manuscript with contributions from other authors. **Competing interests:** A patent application (CN 202210087257.X) has been filed relating to this work. The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text, supplementary materials, or the listed Protein Data Bank (PDB) files. For structures of pRib-ADP bound EDS1-PAD4 and apo-EDS1-PAD4, the atomic coordinates have been deposited in the PDB with accession codes 7XEY and 7XDD, respectively. The EM map for apo-EDS1-PAD4 has been deposited in the Electron Microscopy Database with accession code EMD-33144. Plasmids for in planta expression are available from J.E.P. under a material transfer agreement with Max Planck Institute for Plant Breeding Research.
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq3297
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (*39*–*53*)
Data S1
MDAR Reproducibility Checklist

Submitted 1 April 2022; accepted 22 June 2022
Published online 7 July 2022
10.1126/science.abq3297

Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity

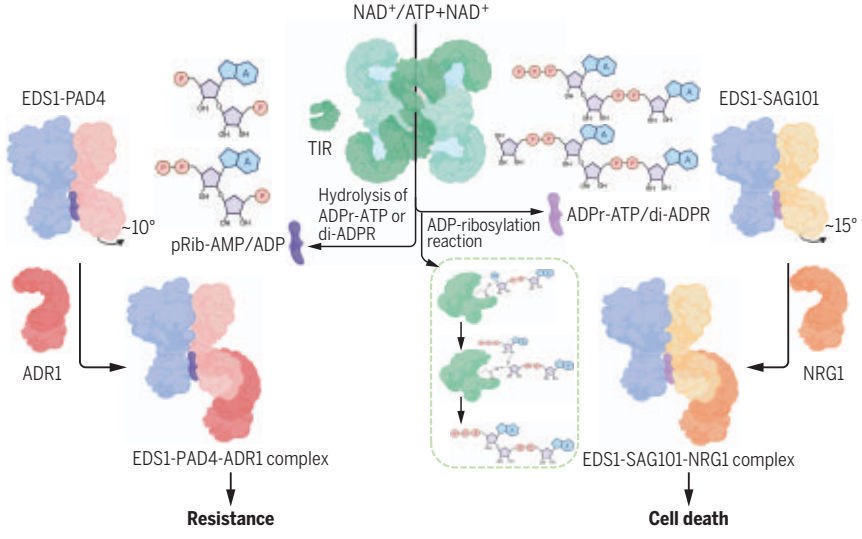
Aolin Jia†, Shijia Huang†, Wen Song†, Junli Wang†, Yonggang Meng†, Yue Sun†, Lina Xu, Henriette Laessle, Jan Jirschitzka, Jiao Hou, Tiantian Zhang, Wenquan Yu, Giuliana Hessler, Ertong Li, Shoucai Ma, Dongli Yu, Jan Gebauer, Ulrich Baumann, Xiaohui Liu, Zhifu Han*, Junbiao Chang*, Jane E. Parker*, Jijie Chai*

INTRODUCTION: Plants deploy intracellular nucleotide-binding leucine-rich repeat (NLR) receptors to detect pathogen effectors that are delivered to host cells during infection. Effector recognition leads to NLR oligomerization, which induces effector-triggered immunity (ETI), often involving host cell death. The NLR receptor subclass called TNL (TIR-NLR) has an N-terminal Toll/interleukin-1 receptor (TIR) signaling domain. Pathogen effector–activated TNLs form tetrameric complexes (resistosomes) with nicotinamide adenine dinucleotide hydrolase (NADase) activity encoded in the TIR domain. The NADase activity of TNLs or TIR domain proteins confers pathogen immunity and/or host cell death. Activated TNLs signal through conserved lipase-like proteins consisting of EDS1 (Enhanced Disease Susceptibility 1) and its two exclusive partners, PAD4 (Phytoalexin Deficient 4) and SAG101 (Senescence-Associated Gene 101), together with a small group of conserved coiled-coil domain–containing helper (signaling) NLRs. In *Arabidopsis*, EDS1-PAD4 and EDS1-SAG101 dimers cooperate with particular helper NLR subgroups, ADR1 (Activated Disease Resistance 1) and NRG1 (N requirement gene 1), respectively, to induce immune responses. The biochemical mechanisms underlying TNL and TIR dependence on these two EDS1 dimer–helper NLR modules remain unknown.

RATIONALE: In *Arabidopsis* TNL- or TIR-triggered immunity, EDS1-PAD4 dimers associate with ADR1-type helper NLRs to restrict pathogen growth, whereas EDS1-SAG101 dimers interact with NRG1-type helper NLRs to promote host cell death. Plant TNLs and TIRs catalyze production of several nucleotide-based molecules in vivo, which suggests that TIR-catalyzed products might activate immune outputs of ADR1 and NRG1. Based on similar but nonidentical EDS1-PAD4 and EDS1-SAG101 surface grooves, we hypothesized that EDS1 dimer binding of TIR NADase–catalyzed products induces association with their corresponding helper NLRs. We identified 2′-(5″-phosphoribosyl)-5′-adenosine diphosphate (pRib-ADP) and monophosphate (pRib-AMP) as the TIR-catalyzed bioactive compounds that bind to and induce EDS1-PAD4 interaction with ADR1. However, these molecules have only weak EDS1-SAG101 binding activity, which suggests that different TIR catalytic products activate the EDS1-SAG101-NRG1 immunity branch.

RESULTS: We found that coexpression of an *Arabidopsis* TNL (RPP1) resistosome or the monocot TIR-only protein from *Brachypodium distachyon* with EDS1, SAG101, and NRG1 induced TNL or TIR NADase–dependent specific interaction between EDS1-SAG101 and NRG1. Coupled with high-resolution mass spectrometry (HRMS) data, a cryo–electron microscopy–generated structure of TNL-activated EDS1-SAG101 revealed that a small molecule, ADP-ribosylated adenosine triphosphate (ADPr-ATP), binds at a similar pocket as pRib-ADP and pRib-AMP to EDS1-PAD4, establishing EDS1-SAG101 as a receptor for this small molecule. ADPr-ATP binding to EDS1-SAG101 induces a conformational change in the C-terminal part of SAG101, which allosterically enables its interaction with NRG1. This mechanism is conserved in pRib-ADP– and pRib-AMP–triggered EDS1-PAD4 binding to ADR1 and explains the recruitment of helper NLR types by their corresponding EDS1 heterodimers. Residues coordinating small molecule binding in both dimers are conserved in seed plant species, suggesting broad relevance. TIR activation resulted in TIR NADase–dependent accumulation of ADPr-ATP in plant tissues. ADPr-ATP is synthesized by TIR-catalyzed transfer of ADP-ribose (ADPR) from NAD^+ (called ADP-ribosylation) to ATP. A related product, ADPr-ADPR (di-ADPR), with similar activity in inducing EDS1-SAG101 interaction with NRG1 is formed by ADP-ribosylation of ADPR. Synthesis of pRib-ADP and pRib-AMP likely involves a two-step mechanism through TIR-catalyzed hydrolysis of ADPr-ATP and di-ADPR.

CONCLUSION: TIR enzyme activity catalyzes ADP-ribosylation of ATP and ADPR to produce NAD^+-derived small molecules that activate two distinctive EDS1 dimer–helper NLR immunity modules. Allosteric activation enables EDS1 dimer association with its cofunctioning helper NLR. The ligands and their receptor mechanisms are likely conserved across seed plants to regulate immune responses. ■



TIR-catalyzed small molecules controlling two immunity branches. Activated TIRs and TNLs use NAD^+ or NAD^+ with ATP as substrates to produce ADPr-ATP and di-ADPR through ADP-ribosylation reactions, which are likely to be further hydrolyzed to pRib-ADP and pRib-AMP. pRib-ADP and pRib-AMP and ADPr-ATP and di-ADPR bind specifically to EDS1-PAD4 and EDS1-SAG101 dimers, triggering conformational changes of PAD4 and SAG101 EP domains to allosterically induce interaction with CNL-type helper NLRs, ADR1 and NRG1, for plant resistance and cell death, respectively.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: chaijj@tsinghua.edu.cn (J.Chai); parker@mpipz.mpg.de (J.E.P.); changjunbiao@zzu.edu.cn (J.Chang); hanzhifu@mail.tsinghua.edu.cn (Z.H.)
†These authors contributed equally to this work.
Cite this article as A. Jia *et al.*, *Science* **377**, eabq8180 (2022). DOI: 10.1126/science.abq8180

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq8180

RESEARCH ARTICLE

PLANT SCIENCE

TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity

Aolin Jia[1]†, Shijia Huang[1]†, Wen Song[2,3]†, Junli Wang[3]†, Yonggang Meng[4]†, Yue Sun[1]†, Lina Xu[5], Henriette Laessle[3], Jan Jirschitzka[2,3], Jiao Hou[6], Tiantian Zhang[6], Wenquan Yu[6], Giuliana Hessler[3], Ertong Li[2], Shoucai Ma[1], Dongli Yu[2,3], Jan Gebauer[2], Ulrich Baumann[2], Xiaohui Liu[5], Zhifu Han[1]*, Junbiao Chang[4,6,7]*, Jane E. Parker[3]*, Jijie Chai[1,2,3]*

Plant pathogen–activated immune signaling by nucleotide-binding leucine-rich repeat (NLR) receptors with an N-terminal Toll/interleukin-1 receptor (TIR) domain converges on Enhanced Disease Susceptibility 1 (EDS1) and its direct partners, Phytoalexin Deficient 4 (PAD4) or Senescence-Associated Gene 101 (SAG101). TIR-encoded nicotinamide adenine dinucleotide hydrolase (NADase) produces signaling molecules to promote exclusive EDS1-PAD4 and EDS1-SAG101 interactions with helper NLR subclasses. In this work, we show that TIR-containing proteins catalyze adenosine diphosphate (ADP)–ribosylation of adenosine triphosphate (ATP) and ADP ribose (ADPR) through ADPR polymerase–like and NADase activity, forming ADP-ribosylated ATP (ADPr-ATP) and ADPr-ADPR (di-ADPR), respectively. Specific binding of ADPr-ATP or di-ADPR allosterically promotes EDS1-SAG101 interaction with helper NLR N requirement gene 1A (NRG1A) in vitro and in planta. Our data reveal an enzymatic activity of TIRs that enables specific activation of the EDS1-SAG101-NRG1 immunity branch.

Cell-surface and intracellular immune receptors detect pathogen disturbance and initiate signaling cascades, which lead to host-mediated defenses (*1*, *2*). The Toll/interleukin-1 receptor (TIR) domain is an immune signaling module in many species (*2*). Bacterial and plant TIRs and human sterile alpha and TIR motif containing 1 (hSARM1) are nicotinamide adenine dinucleotide (NAD^+)–hydrolyzing enzymes (*3*–*5*). In plants, intracellular nucleotide-binding leucine-rich repeat (NLR) receptors recognize pathogen-delivered effectors to confer effector-triggered immunity (ETI) (*6*). ETI transcriptionally boosts generally weaker basal immunity mediated by cell-surface pattern-recognition receptors conferring pattern-triggered immunity (PTI), often leading to localized host cell death (*7*, *8*). A major pathogen-sensing NLR subclass (called TIR-NLRs or TNLs) has N-terminal TIR signaling domains (*2*, *5*). Also, TIR domain proteins, which lack leucine-rich repeats (TIRs), contribute to ETI and basal immunity responses (*5*, *9*–*11*). The signaling role of TNLs and TIRs depends on their NAD^+ hydrolase (NADase) activity (*3*, *4*, *12*), which is promoted by effector-induced TNL oligomerization (*12*, *13*) or TIR self-association (*3*, *4*, *14*).

Plant immunity mediated by TNLs and TIRs requires the conserved lipase-like protein Enhanced Disease Susceptibility 1 (EDS1) and its sequence-related direct partners, Phytoalexin Deficient 4 (PAD4) and Senescence-Associated Gene 101 (SAG101) (*15*–*17*), and two subfamilies of coiled-coil (CC)–like N-terminal domain helper NLRs, Activated Disease Resistance 1 (ADR1) and N requirement gene 1 (NRG1) (*16*, *18*–*22*). Studies in *Arabidopsis* and *Nicotiana benthamiana* (*Nb* tobacco) have revealed that EDS1-SAG101 cooperates with NRG1-type helper NLRs as a TNL ETI-induced module promoting host cell death (*16*, *23*). Functional cooperation and pathogen-induced interaction between the *Arabidopsis* EDS1-PAD4 dimer and ADR1-type helper NLRs, by contrast, is associated with boosted basal immunity and restriction of pathogen growth (*16*, *18*, *23*–*26*). In *Arabidopsis*, these two modules are partially redundant in their functions, but they also display preferences for specific immune outputs (*27*).

TIR-catalyzed small molecules induced the assembly of *Arabidopsis* EDS1-SAG101-NRG1 and EDS1-PAD4-ADR1 complexes in insect cells (*28*). TNL- and TIR-generated 2′-(5′′-phosphoribosyl)-5′-adenosine monophosphate (pRib-AMP) and diphosphate (pRib-ADP) efficiently induced EDS1-PAD4 association with ADR1-L1 but displayed only weak activity in promoting EDS1-SAG101 interaction with NRG1A (*28*). These data and the EDS1-SAG101-NRG1 signaling preferences in vivo suggested that other, yet-unidentified TNL- or TIR-catalyzed small molecules underlie signaling through this immune complex. We show that TIR domain proteins catalyze an ADP ribose-transfer reaction using NAD^+ or NAD^+ with ATP as substrates to produce ADP-ribosylated ADPR (di-ADPR) and ADP-ribosylated ATP (ADPr-ATP), respectively. Like pRib-ADP and pRib-AMP (*28*), di-ADPR and ADPr-ATP can function as second messengers to induce TNL signaling. The latter two molecules specifically bind to and induce *Arabidopsis* EDS1-SAG101 interaction with NRG1A, establishing EDS1-SAG101 as a receptor complex for di-ADPR and ADPr-ATP. Structural analyses reveal specific recognition mechanisms for di-ADPR and ADPr-ATP by EDS1-SAG101. Our data show that different small molecules catalyzed by TIR proteins selectively potentiate two distinctive EDS1-mediated signaling complexes and shed light on the catalytic mechanism underlying production of these small molecules.

Identification of a small molecule specifically activating EDS1-SAG101-NRG1

pRib-ADP and pRib-AMP displayed weak activity in inducing the EDS1-SAG101 interaction with NRG1A (*28*), which suggests that other TIR-catalyzed small molecules promote this interaction more efficiently. When coexpressed with the *Arabidopsis* TNL receptor RPP1 and its cognate *Hyaloperonospora arabidopsidis* effector ATR1 (*29*), EDS1-SAG101 was copurified with N-terminally GST-fused NRG1A (fig. S1A) (*28*). NRG1A binding of EDS1-SAG101 was suppressed by mutation of RPP1 catalytic residue Glu^{158}, which suggests that the EDS1-SAG101 interaction with NRG1A is conferred by RPP1-catalyzed products. We denatured the EDS1-SAG101 protein coexpressed with RPP1 resistosome and extracted EDS1-SAG101–bound small molecules. Although separately purified apo-EDS1-SAG101 failed to interact with NRG1A, the complex gained NRG1A- but not ADR1-L1–binding activity when incubated with the extractant (fig. S1B). Liquid chromatography and high-resolution mass spectrometry (LC-HRMS) analysis of the extractant revealed a chemical species with a mass-to-charge ratio (m/z) of 1047.0530 ($z = 1^-$) (Fig. 1A). Additionally, a trace amount of a product with a molecular weight matching that of pRib-ADP was detected (fig. S1C), supporting weak pRib-ADP activity in promoting EDS1-SAG101 association with NRG1A (*28*).

To identify the chemical species with m/z 1047.0530 ($z = 1^-$), we solved a cryo–electron microscopy (cryo-EM) structure of EDS1-SAG101 purified after coexpression with RPP1 and ATR1 in insect cells (Fig. 1B, fig. S2, and table S1). A cryo-EM density not occupied by EDS1-SAG101

[1]Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, Center for Plant Biology, School of Life Sciences, Tsinghua University, 100084 Beijing, China. [2]Institute of Biochemistry, University of Cologne, 50674 Cologne, Germany. [3]Department of Plant-Microbe Interactions, Max-Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. [4]School of Pharmaceutical Sciences, Zhengzhou University, 450001 Zhengzhou, China. [5]National Protein Science Facility, Tsinghua University, 100084 Beijing, China. [6]College of Chemistry, Zhengzhou University, 450001 Zhengzhou, China. [7]Henan Key Laboratory of Organic Functional Molecules and Drug Innovation, Henan Normal University, 453007 Xinxiang, China.
*Corresponding author. Email: chaijj@tsinghua.edu.cn (J.Chai); parker@mpipz.mpg.de (J.E.P.); changjunbiao@zzu.edu.cn (J.Chang); hanzhifu@mail.tsinghua.edu.cn (Z.H.)
†These authors contributed equally to this work.

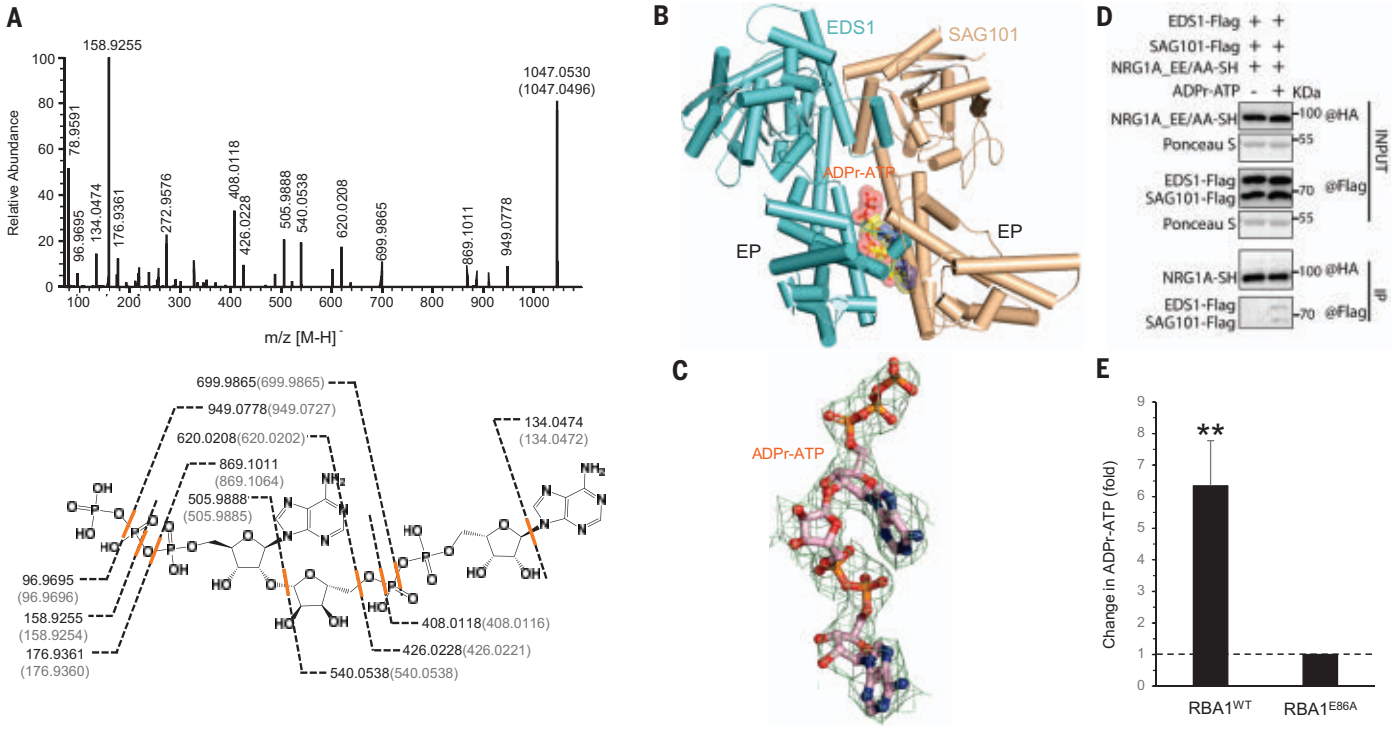


Fig. 1. Identification of a TIR-catalyzed small molecule activating EDS1-SAG101. (**A**) Mass analysis of the small molecule bound to EDS1-SAG101. (Top) MS/MS spectra after collision-induced dissociation (CID) fragmentation of the ion with *m/z* 1047.0530 ($z = 1^-$). (Bottom) Scheme showing the proposed routes of formation of various ions shown above. Numbers in parentheses indicate theoretical molecular weights of proposed ions. (**B**) Cryo-EM structure of EDS1-SAG101 bound by ADPr-ATP. ADPr-ATP bound between EDS1 and SAG101 EP domains is shown in transparent sphere. EDS1 and SAG101 are shown in cartoon, and ADPr-ATP is shown in stick. (**C**) Cryo-EM density around the bound ADPr-ATP (in stick). The map was contoured at 8.0 sigma in Pymol. (**D**) ADPr-ATP induces EDS1-SAG101 interaction with NRG1A after protein expression in *Nb* tobacco. ADPr-ATP previously extracted from EDS1-SAG101 was incubated with a lysate of *Nb* tobacco quadruple mutant *eds1pad4sag101ab* (*Nb-epss*) expressing EDS1-FLAG, SAG101-FLAG, and NRG1A$^{E14A/E27A}$-SH at 4°C for 1 hour. Protein-protein interaction was detected by Co-IP with antibodies as indicated. (**E**) Relative change in ADPr-ATP accumulation in *Nb-epss* plants expressing RBA1^{WT} or RBA1^{E86A}. The statistics are based on LC-MS results of three replicates. The amount of ADPr-ATP in RBA1^{E86A}-expressing *Nb-epss* was normalized to 1.0. Significance calculated with Tukey's HSD (honestly significant difference) test ($n = 3$, $\alpha = 0.05$; different letters indicate significant differences). *Nb-epss* leaves expressing *RBA1*^{WT} or the catalytic mutant *RBA1*^{E86A} and insect cells expressing apo-EDS1-SAG101 were mixed and cosonicated. The apo-EDS1-SAG101 complex in the lysate was purified through affinity chromatography. Multiple reaction monitoring (MRM) analyses were performed for ADPr-ATP bound in the purified EDS1-SAG101 complex. Significance of the difference was determined using Student's *t* test ($n = 3$; $^{**}P < 0.01$).

is clearly discerned at the interface between the C-terminal EP domains of EDS1 and SAG101 (Fig. 1C), corresponding to where pRib-ADP binds in EDS1-PAD4 (fig. S3A) (*28*). pRib-ADP fitted well into half of the unoccupied cryo-EM density (fig. S3B), but two patches of cryo-EM density remained unfilled after pRib-ADP modeling (fig. S3B). The smaller patch, which connects to the terminal (β) phosphate group of the modeled pRib-ADP, likely results from a phosphate group. The other patch is much larger and accommodates an AMP molecule. The phosphate group, pRib-ADP, and pRib-AMP are subgroups of ADPr-ATP, with a calculated molecular mass of 1048.0496, matching 1048.0530 from the LC-HRMS analysis (Fig. 1A, top). Furthermore, prominent ions with *m/z* values expected for product ions of ADPr-ATP were observed in the high-resolution tandem mass spectrometry (MS/MS) spectrum (Fig. 1A, bottom).

We investigated whether ADPr-ATP extracted and purified as described above induces an EDS1-SAG101-NRG1A complex when expressed in *Nb* tobacco. *EDS1*, *SAG101*, and an *NRG1A* E14A/E27A cell death–inactive variant (*23*, *30*, *31*) were transiently coexpressed in an *Nb* tobacco *epss* line lacking all EDS1-family proteins (*Nb-epss*) (*16*). ADPr-ATP was incubated with the *Nb-epss* tissue lysate. Incubation of ADPr-ATP resulted in an interaction between EDS1-SAG101 and NRG1A, which was not detected in the absence of ADPr-ATP (Fig. 1D). These results indicate that ADPr-ATP is sufficient to promote EDS1-SAG101 association with NRG1A.

We next used the *Nb-epss* transient assay to investigate whether a TIR protein catalyzes accumulation of ADPr-ATP in plants. *Arabidopsis* RBA1 (*9*) was used for this because of its *EDS1*-dependent cell death activity in *Nb* tobacco (*28*). *Nb-epss* leaves expressing RBA1 were cosonicated with insect cells expressing EDS1 and SAG101, and LC-MS was used to detect small molecules captured by purified EDS1-SAG101. As shown in Fig. 1E, plants expressing wild-type RBA1 but not an RBA1 catalytic mutant (E86A) produced strong ADPr-ATP signals in the assay. These data show that TIR-catalyzed ADPr-ATP is produced in plant tissues and binds to EDS1-SAG101.

Recognition mechanism of ADPr-ATP by EDS1-SAG101

ADPr-ATP adopts an elongated conformation binding to a positively charged pocket between the EDS1 and SAG101 EP domains (Fig. 2A). The pRib-ADP moiety in ADPr-ATP forms a large network of hydrogen bonding and polar interactions with EDS1-SAG101 (Fig. 2B) that are conserved in the structure of pRib-ADP–bound EDS1-PAD4 (fig. S3C). The remaining portions of ADPr-ATP also establish extensive contacts with EDS1-SAG101. The terminal nucleobase adenine hydrophobically packs against SAG101 Met^{384} and Leu^{438} at one side and EDS1 His^{476} at the other side of the pocket (Fig. 2B). The terminal phosphate group is solvent

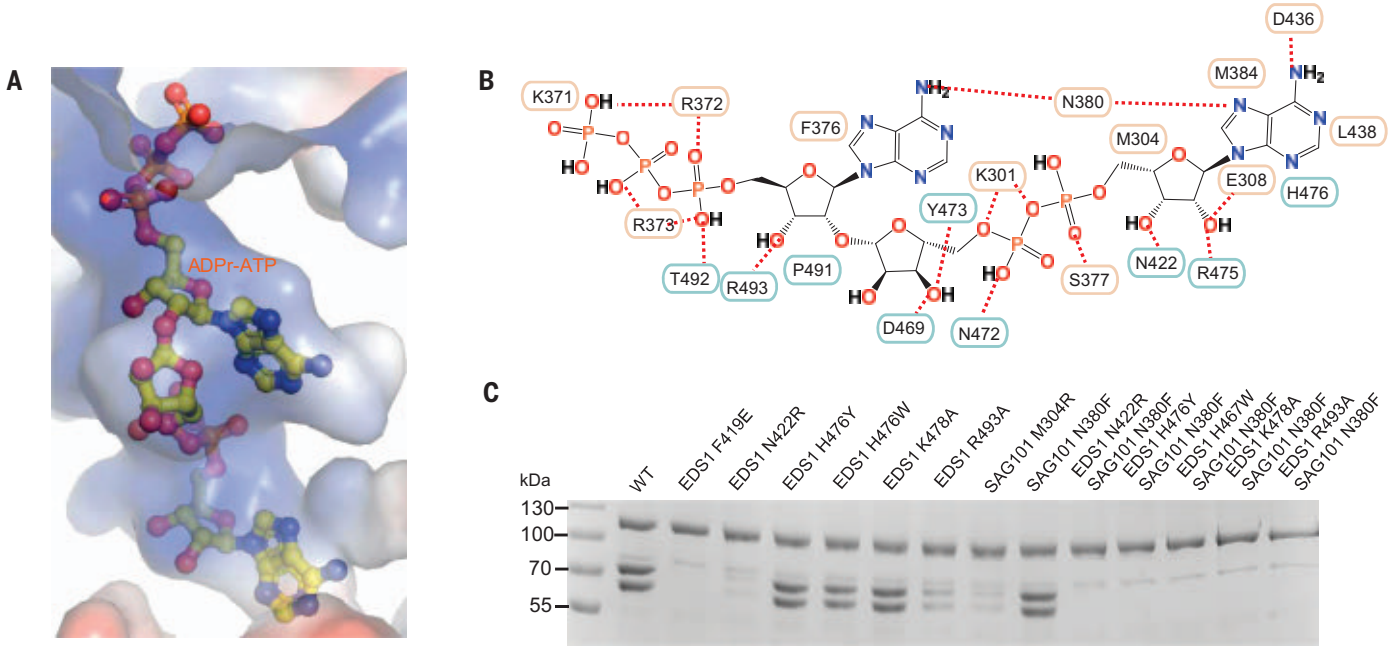


Fig. 2. ADPr-ATP–specific recognition by EDS1-SAG101. (**A**) Close-up view of the ADPr-ATP binding pocket of EDS1-SAG101. EDS1-SAG101 is shown in electrostatic surface, and ADPr-ATP is shown in stick. Blue, red, and white represent positive, negative, and neutral surfaces, respectively. (**B**) Diagram showing coordinates for interaction between EDS1-SAG101 and ADPr-ATP. Cyan and purple frames represent residues from EDS1 and SAG101, respectively. Red dashed lines represent polar interactions. (**C**) Analysis of EDS1-SAG101 ADPr-ATP binding pocket mutants for induced association with NRG1A. N-terminally GST-NRG1A was coexpressed with nontagged EDS1 and C-terminally Strep-tagged SAG101 (wild-type or mutant forms as indicated), RPP1, and 10×His-tagged ATR1 in insect cells. GST-NRG1A was purified using GS4B beads. GS4B-bound proteins were eluted, separated by SDS–polyacrylamide gel electrophoresis (SDS-PAGE), and detected by Coomassie brilliant blue staining. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

exposed and only makes van der Waals contact with SAG101 Lys^{371}. Sequence alignment showed that the ADPr-ATP–interacting residues of EDS1 (fig. S4A) and SAG101 (fig. S4B) are conserved among dicot plant species.

To verify the ADPr-ATP–activated EDS1-SAG101 cryo-EM structure, we assessed the impact of mutations at the ADPr-ATP binding pocket on EDS1-SAG101 interaction with NRG1A using the insect cell expression system and pull-down assay described above. Substitution of SAG101 Met^{304} with its PAD4 equivalent Arg^{314} substantially weakened EDS1-SAG101 interaction with NRG1A (Fig. 2C), which explains why ADPr-ATP was barely active in stimulating EDS1-PAD4 interaction with ADR1-L1 (fig. S1B). In *Arabidopsis*, SAG101 Met^{304} and PAD4 Arg^{314} were essential for signaling through their respective branches and induced association with corresponding helper NLRs (*32*). By contrast, the interaction was only slightly affected by substitution of SAG101 Asn^{380} with its PAD4 equivalent Phe^{387} (Fig. 2C). Supporting a specific role for EDS1 His^{476} in ADPr-ATP recognition, H476Y substantially suppressed AvrRps4-induced cell death but had little effect on AvrRps4-induced immunity in *Arabidopsis* (*16*). Although the EDS1 H476Y mutation did not disable EDS1-SAG101 interaction with NRG1A (Fig. 2C), combination of SAG101 N380F and EDS1 H476Y resulted in a loss of NRG1A-binding activity of EDS1-SAG101. An R493A mutation in EDS1 only reduced EDS1-SAG101 interaction with NRG1A (Fig. 2C), in contrast with a loss of pRib-ADP–induced EDS1-PAD4 interaction with ADR1-L1 (*28*). This would explain why the EDS1 mutation disturbed mainly EDS1-PAD4–mediated pathogen resistance in vivo (*16*, *25*). The difference might be the result of more extensive contacts between ADPr-ATP and EDS1-SAG101, rendering binding of the small molecule less sensitive to mutation at this position.

Selective EDS1-SAG101 activation mechanism by ADPr-ATP

ADPr-ATP specifically promotes EDS1-SAG101 interaction with NRG1A (fig. S1B). In comparison, pRib-ADP induces EDS1 heterodimer interactions with ADR1-L1 and NRG1A, although with different efficiency (*28*). pRib-ADP and ADPr-ATP bind to a similar position in the EDS1-PAD4 and EDS1-SAG101 complexes (fig. S3A). The pRib-ADP–contacting residues of PAD4 are largely conserved in SAG101 (fig. S3C), whereas SAG101 residues for interacting with the ADPR moiety of ADPr-ATP are largely variable in PAD4 (fig. S5). Notably, substitutions of SAG101 Met^{304} and Asn^{380} with their bulkier equivalents PAD4 Arg^{314} and Phe^{387} result in occlusion of the ADPR-binding site in EDS1-PAD4 (Fig. 3A). The different rotamers of PAD4 Gln^{318} and its counterpart SAG101 Glu^{308} further sterically block ADPr-ATP from being recognized by EDS1-PAD4. These structural observations provide an explanation for specific recognition of ADPr-ATP by EDS1-SAG101.

Superposition of an apo-EDS1-SAG101 crystal structure (*15*) with the cryo-EM structure of ADPr-ATP–bound EDS1-SAG101 revealed that the EDS1 conformation is nearly identical in the two forms of EDS1-SAG101 (Fig. 3B). There are notable structural differences in the SAG101 EP domain but only a slight conformational change in the SAG101 lipase-like domain upon ADPr-ATP binding (Fig. 3B). Compared with the apo form of SAG101, the EP domain in ADPr-ATP–bound SAG101 rotates anticlockwise ~15 degrees (viewed from the bottom) (Fig. 3C). This causes complete melting of α16 and shifts of the N-terminal sides of helices α14 and α15 toward EDS1, whereas SAG101 α11 shifts ~4.0 Å away from EDS1. These structural analyses indicate that ADPr-ATP binding triggers conformational changes in the SAG101 EP domain which enable EDS1-SAG101 interaction with NRG1A. Like pRib-ADP bound in EDS1-PAD4, ADPr-ATP bound in EDS1-SAG101 is almost completely buried, which indicates

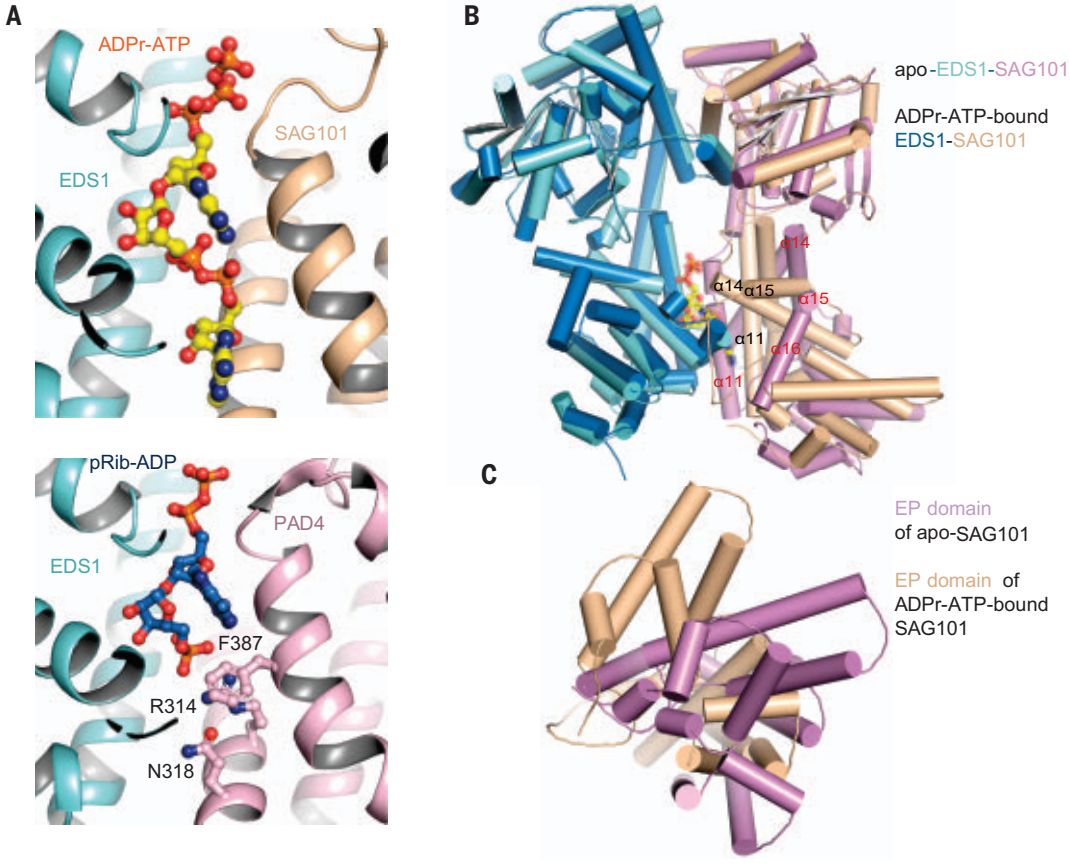


Fig. 3. ADPr-ATP selection and activation mechanism of EDS1-SAG101. (**A**) Structural alignment of EDS1-SAG101 (top) and EDS1-PAD4 (bottom; PDB code: 7XEY) around the ligand-binding sites. EDS1 (cyan), SAG101 (light orange), and PAD4 (pink) are shown in cartoon, and ADPr-ATP (yellow) and pRib-ADP (blue) are shown in stick. Three PAD4 residues, Phe^{387}, Arg^{314}, and Asn^{318}, not conserved in SAG101 are shown. (**B**) Structural alignment of apo-EDS1-SAG101 and ADPr-ATP–bound EDS1-SAG101. Color codes are indicated. Secondary structural elements of SAG101 around the ADPr-ATP–binding site are labeled. (**C**) Bottom view of the EP domain structural alignment between apo-EDS1-SAG101 and ADPr-ATP–bound EDS1-SAG101 shown in (B).

that this small molecule is unlikely to participate in direct interaction with NRG1A. Thus, ADPr-ATP and pRib-ADP (*28*) use a conserved allosteric mechanism to activate EDS1-SAG101 and EDS1-PAD4, respectively.

TIR-catalyzed di-ADPR induces EDS1-SAG101 interaction with NRG1A

Expression of RBA1 resulted in accumulation of ADPr-ATP in *Nb* tobacco (Fig. 1E). We next investigated whether a purified TIR protein catalyzes production of ADPr-ATP when NAD^+ is used as a substrate. An isolated RPS4 TNL receptor TIR domain (residues 1 to 236; RPS4-TIR), which causes hypersensitive response (HR)–like cell death when expressed in planta (*33*), was used because it was easily purified. As shown in Fig. 4A, incubation of EDS1-SAG101 with wild-type RPS4-TIR but not its catalytic mutant E88A in the presence of NAD^+ induced EDS1-SAG101 interaction with NRG1A. EDS1-SAG101 displayed no NRG1A-binding activity in the absence of NAD^+. These results suggest that RPS4-TIR–catalyzed ADPr-ATP production supports EDS1-SAG101 interaction with NRG1A. However, LC-HRMS analysis of extracted chemical compounds from the EDS1-SAG101 protein incubated with RPS4-TIR and NAD^+ revealed a strong peak at m/z 1099.1276 ($z = 1^-$) (Fig. 4B) instead of the predicted m/z 1047.0530 ($z = 1^-$) of ADPr-ATP. Given its EDS1-SAG101–binding activity, the unknown substance with m/z 1099.1276 ($z = 1^-$) should be structurally related to ADPr-ATP. One such candidate is di-ADPR with molecular weight (1100.1255), which matches the 1100.1276 peak measured by LC-HRMS. Supporting this assignment, prominent ions with m/z values expected for product ions of di-ADPR were observed in the high-resolution MS/MS spectrum (Fig. 4C).

The activity of ADPr-ATP and di-ADPR in inducing EDS1-SAG101 interaction with NRG1A suggests that the terminal ribose sugar ring of di-ADPR or the γ phosphate of ADPr-ATP is not essential for EDS1-SAG101 binding. This is consistent with the observation that the γ phosphate is solvent exposed in the cryo-EM structure of EDS1-SAG101 (Fig. 1B). Further supporting this conclusion, a chemically synthesized di-ADPR enantiomer with the configuration of C-1″ from the terminal ribose sugar ring inverted (fig. S6 and data S1) induced the EDS1 heterodimer interaction with NRG1A (Fig. 4D). By contrast, the synthetic compound had little activity of potentiating EDS1-PAD4 binding to ADR1-L1 in the assay (Fig. 4D).

Mechanism of TIR-catalyzed production of bioactive small molecules

TIR-catalyzed production of ADPr-ATP seemingly results from condensation between ATP and ADPR. However, incubation of RPS4-TIR, ATP, and ADPR failed to promote EDS1-SAG101 interaction with NRG1A (Fig. 4A). This result indicates that neither ADPr-ATP nor di-ADPR was produced in the reaction, and ADPR is unlikely to be the sole RPS4-TIR substrate for di-ADPR production. Given that NADase activity is required for TIR-induced EDS1-SAG101 interaction with NRG1A (Fig. 4A and fig. S1A), we reasoned that formation of the 1″-2′ glycoside bond in ADPr-ATP and di-ADPR involves NAD^+ hydrolysis. Incubation of an equal molar ratio of ATP and NAD^+ with RPS4-TIR conferred EDS1-SAG101 NRG1A-binding activity (Fig. 4A). LC-HRMS showed that both di-ADPR and ADPr-ATP were produced in this reaction, with di-ADPR being the major product (fig. S7A). Similar results were obtained when RPS4-TIR was replaced with the RPP1 resistosome purified from insect cells (fig. S8). However, ADPr-ATP was predominantly captured by EDS1-SAG101 (fig. S7B), which suggests a higher EDS1-SAG101 affinity for ADPr-ATP than for di-ADPR.

A crystal structure of RPP1-TIR (residues 83 to 254) solved in the presence of NAD^+ revealed that the TIR domains form a tetramer that is nearly identical to that in the cryo-EM structure of the RPP1 resistosome (fig. S9 and table S2) (*12*). Well-resolved electron densities at the two asymmetric dimer interfaces (Fig. 5A), where the ATP molecules are bound in the RPP1 (*12*) and

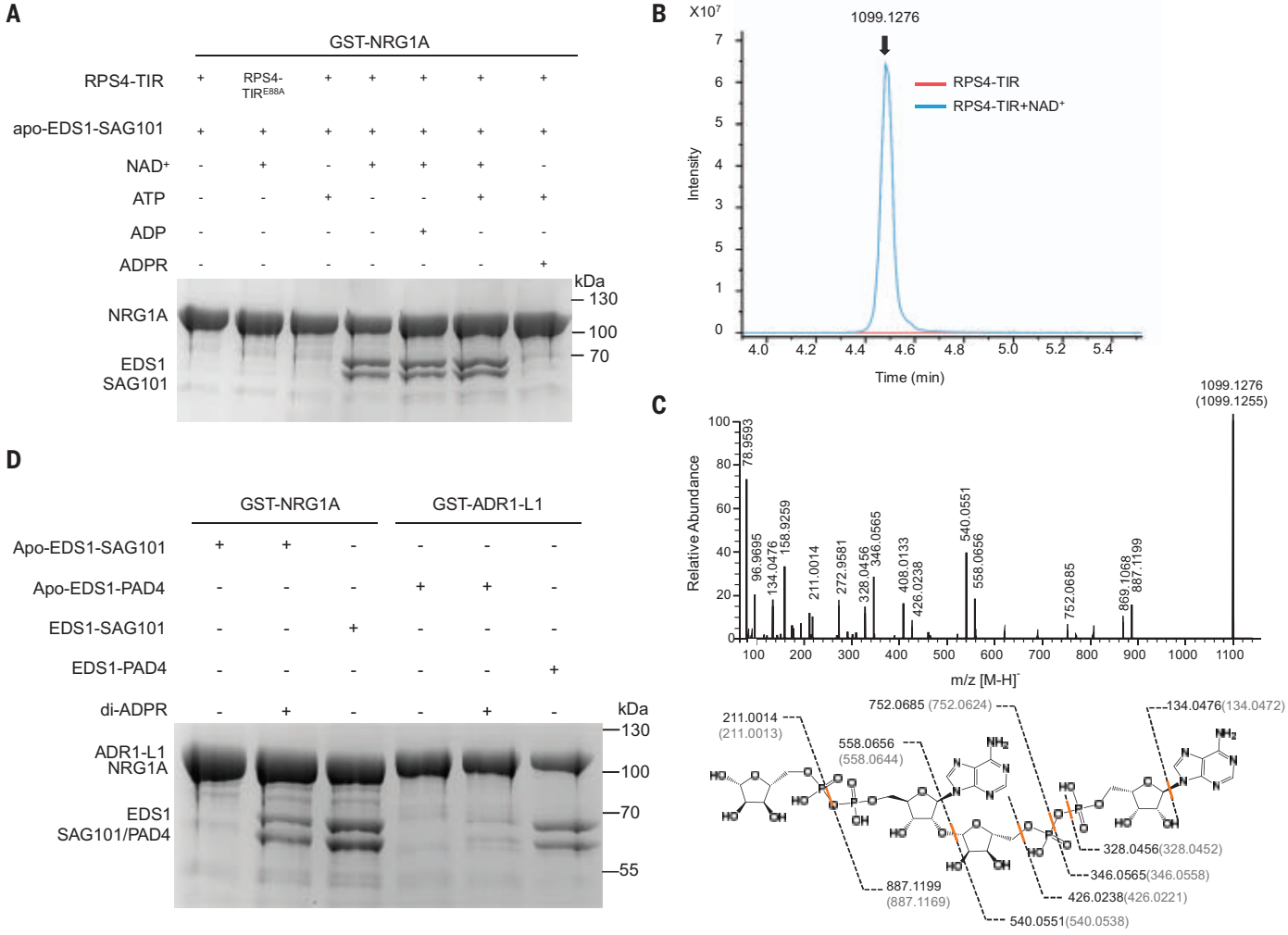


Fig. 4. TIR-catalyzed di-ADPR induces EDS1-SAG101 interaction with NRG1A. (**A**) RPS4 TIR catalyzes different substrates to produce small molecules, inducing EDS1-SAG101 interaction with NRG1A. Reaction products of RPS4 TIR domain (residues 1 to 236) protein with substrates indicated were incubated with GS4B beads bound by GST-tagged NRG1A preincubated with excess apo-EDS1-SAG101 on ice for 30 min. After washing, proteins bound in GS4B beads were eluted, analyzed by SDS-PAGE, and detected by Coomassie brilliant blue staining. (**B**) LC-HRMS chromatograms of EDS1-SAG101–captured small molecules from RPS4-TIR–catalyzed products with NAD^+ as the substrate. Purified RPS4-TIR protein was incubated with NAD^+ at room temperature for 16 hours. Products of the reaction were captured by apo-EDS1-SAG101, and the captured small molecules were extracted by denaturing the complex with 80% methanol (v/v) and then subjected to LC-HRMS. (**C**) Mass analysis of small molecules in (B). (Top) MS/MS spectra after CID fragmentation of the ion with *m/z* 1099.1276 ($z = 1^-$). (Bottom) Scheme showing proposed routes of formation of various ions shown above. Numbers in parentheses indicate theoretical molecular weights of proposed ions. (**D**) A synthetic di-ADPR enantiomer induces EDS1-SAG101 interaction with NRG1A. The synthetic compound was incubated with GS4B beads bound by GST-tagged NRG1A and ADR1-L1 preincubated with excess apo-EDS1-SAG101 and apo-EDS1-PAD4 on ice for 30 min. After washing, proteins bound in GS4B beads were eluted, analyzed by SDS-PAGE, and detected by Coomassie brilliant blue staining.

ROQ1 (*13*) resistosomes (fig. S9A), match well with ADPR, which indicates that the crystal structure represents an RPP1-TIR product complex. C1″ in the ribose sugar ring of ADPR faces the catalytic residue Glu^{158} (Fig. 5B), suggesting that this residue is likely important for catalyzing nicotinamide release from NAD^+. ADPR binding to the asymmetric RPP1 TIR dimer is markedly similar to that of a substrate mimic to the active site of hSARM1 (fig. S9B) (*34*). These data suggest that formation of the 1″-2′ glycoside bond in ADPr-ATP involves nucleophilic attack on the C1″ of the TIR-activated NAD^+ by the 2′-OH of ATP, resulting in ADP-ribosylation of the in-coming ATP molecule. ADPR is known to be a product of TIR-catalyzed NAD^+ hydrolysis (*3*, *4*). Thus, di-ADPR is produced through ADP-ribosylation of ADPR.

Our previous study had shown that TIR proteins can hydrolyze ATP and guanosine 5′-triphosphate (GTP) (*35*). We therefore tested whether pRib-ADP and pRib-AMP (which efficiently activate EDS1-PAD4) (*28*) can be produced through RPS4-TIR hydrolysis of ADPr-ATP or di-ADPR (which activate EDS1-SAG101). NAD^+, NAD^+ + ATP, or NAD^+ + ADP were incubated with bacteria-purified RPS4-TIR and insect cell–purified apo-EDS1-PAD4 was used to capture the products of each reaction. The EDS1-PAD4–captured molecules were detected by LC-HRMS. Both pRib-ADP and pRib-AMP were captured by EDS1-PAD4 when ATP + NAD^+ were the substrates (Fig. 5C). By contrast, much lower pRib-ADP amounts were captured when NAD^+ alone was the substrate. Coincubation of NAD^+ with ADP notably also promoted production of pRib-ADP and pRib-AMP. These data point to a two-step reaction in which RPS4-TIR first generates ADPr-ATP and di-ADPR through ADP-ribosylation of ATP and ADPR and then hydrolyzes the products to form pRib-ADP and pRib-AMP (fig. S10).

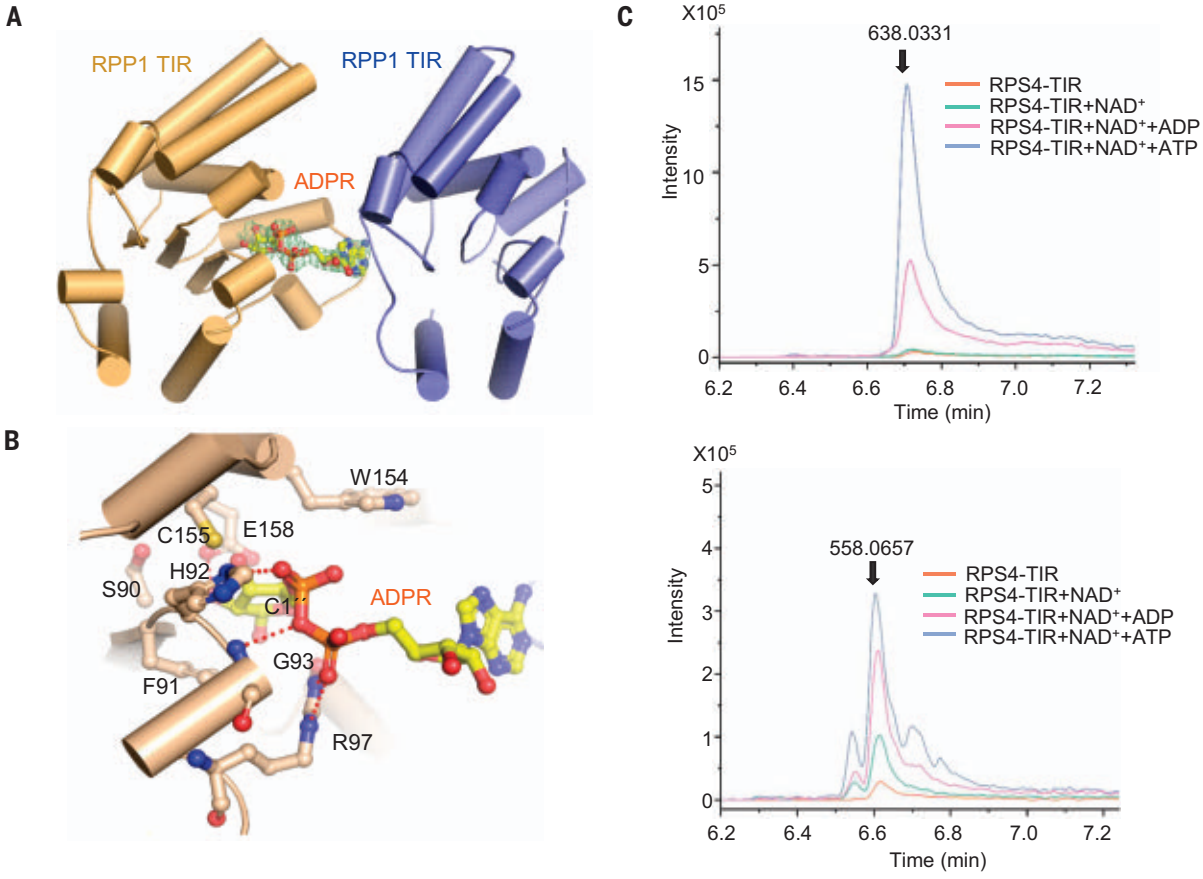


Fig. 5. Mechanism for TIR-catalyzed production of ADPr-ATP, di-ADPR, pRib-AMP, and pRib-ADP. (**A**) An asymmetric RPP1 TIR dimer with ADPR bound. Omit electron density (*2Fo-Fc*) contoured at 1.3 sigma surround ADPR (in stick). For clarity, only one asymmetric RPP1 TIR dimer from a tetrameric RPP1 TIR is shown. (**B**) Detailed interactions between the ribose sugar ring and one RPP1 TIR. Red dashed lines represent polar interactions. The C-1′′ atom is labeled. (**C**) LC-HRMS chromatograms of apo-EDS1-PAD4–captured pRib-ADP (top) and pRib-AMP (bottom) from RPS4-TIR–catalyzed products. Different colors represent different substrates of RPS4-TIR. Purified RPS4-TIR protein was incubated with the indicated substrates at room temperature for 16 hours. Products of each reaction were captured by apo-EDS1-PAD4, and the captured small molecules were extracted by denaturing the complex with 80% methanol (v/v) and subjected to LC-HRMS analyses.

Discussion

In this work and a related study (*28*), we discover a broadly conserved allosteric mechanism for the activation of EDS1-PAD4 and EDS1-SAG101 by TIR-catalyzed small molecules. The induced conformational change in the two EDS1 heterodimers promotes direct association with cofunctioning helper NLRs, presumably to promote their Ca^{2+}-permeable ion channel activity (*30*). Despite the shared biochemical activity, resulting EDS1-PAD4-ADR1 and EDS1-SAG101-NRG1 signaling complexes display preferences for specific immune outputs (*16*, *23*, *32*). This can be partly reconciled by the existence of multiple TIR-catalyzed signaling molecules with overlapping and distinct binding specificities for the two EDS1 heterodimers (*28*) (Fig. 3A). Amino acids that influence the selection of these small molecules by EDS1 heterodimers (Fig. 2 and Fig. 3A) and immunity functions in vivo (*16*, *23*, *25*, *32*) are located at similar EDS1-PAD4 and EDS1-SAG101 EP domain surfaces. Notably, an EDS1 mutation (H476Y) that perturbed the ligand binding site of EDS1-SAG101 but not EDS1-PAD4 (Fig. 2B and Fig. 3A) abolished host cell death but had little effect on resistance to bacteria in *Arabidopsis* (*16*). Thus, specific immune functions of the two EDS1 heterodimers appear to be conferred by their recognized ligands. This is consistent with the observation that overexpression of EDS1 with PAD4 leads to increased *Arabidopsis* pathogen resistance but not host cell death (*24*). The strong activity of pRib-ADP and pRib-AMP in promoting EDS1-PAD4 interaction with ADR1, but a very low capacity for promoting EDS1-SAG101 interaction with NRG1A (*28*), suggests that these TIR products might induce effective plant immunity without cell death. The identified TIR-catalyzed signaling molecules are produced with unequal molar ratios (Fig. 5C and fig. S7A), probably because of differences in their stability (*28*) and production efficiency. In plant tissues, TIR-catalyzed small molecules might be subject to temporal and spatial regulation to ensure that one particular set accumulates. Such regulation might be an important control mechanism in activating ADR1 or NRG1 or both helper NLRs. Future analysis of small molecules and their receptor dynamics can help to develop strategies for the manipulation of plant immunity.

Structure-based sequence alignments indicate that the ligand binding pockets of EDS1-PAD4 and EDS1-SAG101 are evolutionarily conserved in different seed plant species (fig. S4) (*28*). A phylogenetic analysis revealed evidence that EDS1-PAD4 and EDS1-SAG101 have coevolved with ADR1 and NRG1, respectively (*16*). Also, a functional incompatibility was reported between *Arabidopsis* and *Solanaceae* (tomato and *Nb* tobacco) EDS1-SAG101 and NRG1 proteins (*16*, *17*). Our structural studies show that pRib-ADP and pRib-AMP binding to EDS1-PAD4 and ADPr-ATP and di-ADPR binding to EDS1-SAG101 almost completely bury these small molecules and the conserved ligand recognition pockets (Fig. 1,

Fig. 2, and fig. S3) (*28*). These observations indicate that the small molecules do not directly participate in the induced EDS1-PAD4 interaction with ADR1 or the EDS1-SAG101 interaction with NRG1. Rather, an allosteric change induced by small molecule binding in each dimer probably creates a new protein surface for specific helper NLR association. Although structural information on the nature of EDS1-PAD4-ADR1 and EDS1-SAG101-NRG1 complexes is needed, we think that ligand-induced conformational changes in the PAD4 or SAG101 EP domains are the most likely structural determinants for lineage-specific helper NLR interactions.

Expression of a TIR-only protein from the monocot *Brachypodium distachyon* that lacks NRG1 and SAG101 causes cell death in *Nb* tobacco (*3*). Although PAD4 has so far not been found to contribute to TNL pathogen resistance and cell death outputs in *Nb* tobacco (*16*, *17*), these results suggest that the monocot TIR protein (*Bd*TIR) catalyzes the production of signaling molecules that activate both EDS1 heterodimers. This was confirmed by our biochemical assays showing that *Bd*TIR induced *Arabidopsis* EDS1-PAD4 interaction with ADR1-L1 (*28*) and EDS1-SAG101 interaction with NRG1A (fig. S11) in insect cell expression and pull-down assays. We speculate that TIR-catalyzed production of di-ADPR and ADPr-ATP might also be a major synthetic pathway for pRib-ADP and pRib-AMP in monocot plants. However, we cannot rule out other functions of di-ADPR and ADPr-ATP or monocot TIR production of other bioactive nucleotide-based molecules.

The TIR-catalyzed ADPR-transfer reactions are reminiscent of ADPR chain extensions catalyzed by poly–ADPR polymerases (PARPs) (*36*), which use a strictly conserved glutamate residue for the catalysis (*36*). The conserved glutamate residue of TIRs may have a similar role in catalyzing ADP-ribosylation of ATP and ADPR by activating the C-1″ of NAD^+ in the active site (fig. S10). C-1″ of ADPr-ATP in the cryo-EM structure retains the same configuration as that of NAD^+, suggesting that TIR-catalyzed ADP-ribosylation might involve a double-displacement catalytic mechanism, as shown for hSARM1 (*34*). However, it is possible that TIR-catalyzed ADPr-ATP or di-ADPR with an inverted configuration at C-1″ cannot be captured by EDS1-SAG101 because of their deficient or weak EDS1-SAG101–binding activity. Thus, the possibility remains that TIR-catalyzed ADP-ribosylation of ATP and ADPR also involves inversion of the configuration of C-1″ at this position, like with PARP-catalyzed ADP-ribosylation of proteins. Regardless of the catalytic mechanism, ADP-ribosylation of ADPR and ATP involves both NADase and PARP-like activities. In addition to ATP and ADPR, ADP can also act as a substrate for a TIR-catalyzed ADP-ribosylation reaction, albeit with much less efficiency (Fig. 4A and Fig. 5C). This raises the question whether TIRs catalyze ADP-ribosylation of other metabolites. Given the high concentrations of ATP in plant cells, ADPr-ATP is likely the major TIR product for activation of EDS1-SAG101 in the immune response. Biochemical data suggested that TIR-mediated hydrolysis of ADPr-ATP and di-ADPR yields pRib-AMP and pRib-ADP (Fig. 5C and fig. S10, A and B), but it remains possible that TIRs also use other mechanisms for catalytic production of the two compounds in vivo.

In addition to pRib-ADP, pRib-AMP, ADPr-ATP, and di-ADPR, TIRs also catalyze production of adenosine 2′,3′-monophosphate (2′,3′-cAMP) and guanosine 2′,3′-monophosphate (2′,3′-cGMP), which have a critical role in EDS1-dependent signaling (*35*). This is seemingly inconsistent with the biochemical data showing that pRib-AMP and pRib-ADP and ADPr-ATP and di-ADPR are sufficient to induce EDS1 heterodimer interactions with their co-functioning helper NLRs (*28*). However, it should be noted that, in vivo, all components have to be in place for ligands to activate receptor signaling. An example of this is activation of the animal NLR NLRP3, which involves two signals—a priming signal for transcriptional up-regulation of NLRP3 and pro–interleukin-1β (IL-1β) and a second signal for induction of the NLRP3 inflammasome (*37*). Accumulating evidence suggests that pRib-AMP and pRib-ADP and ADPr-ATP and di-ADPR might not be sufficient to induce EDS1-dependent immunity in planta. Estradiol-inducible expression of the pathogen effector AvrRps4 fails to induce cell death in *Nb* tobacco (*38*), and AvrRps4-induced cell death can be potentiated by PTI signaling (*7*, *8*). Supporting a role of PTI in TIR signaling, flg22 (a bacterial PAMP) treatment strongly promotes *Arabidopsis* TIR protein *At*TN3-mediated EDS1-dependent cell death in *Nb* tobacco (*39*). These results suggest that TIR NADase activity is not sufficient to induce EDS1-dependent immunity. RBA1 mutants C83A and KKK^{AAA} lacking the 2′,3′-cAMP and 2′,3′-cGMP synthetase but retaining the NADase activity fail to elicit cell death in *Nb* tobacco (*35*). We found that these two mutants retained activity in promoting EDS1-SAG101 interaction with NRG1A (fig. S12). This result is consistent with the hypothesis that TIR-catalyzed products are insufficient for cell death activation (*14*). TIR domain–containing genes that are up-regulated during early PTI were proposed to produce NADase-catalyzed small molecules to boost PTI signaling (*40*). Whether these TIRs and their catalytic products have a role in PTI potentiation of ETI signaling (*7*, *8*) requires future study. It currently remains unknown how 2′,3′-cAMP and 2′,3′-cGMP cooperate with pRib-ADP, pRib-AMP, ADPr-ATP, and di-ADPR to potentiate EDS1-dependent immune responses. We speculate that 2′,3′-cAMP and 2′,3′-cGMP may up-regulate the expression of genes required for EDS1 signaling. This model agrees with a recent study showing that 2′,3′-cAMP induces many stress-related *Arabidopsis* genes (*41*).

Materials and methods summary

Nontagged EDS1 and C-terminally twin-StrepII tagged SAG101 were coexpressed in Sf21 cells, and the apo-EDS1-SAG101 complex was purified using Streptactin beads. The EDS1 and SAG101 constructs were also used to co-express with RPP1_WsB (C-terminally twin-StrepII tagged) and ATR1_Emoy2 or RPS4-TIR (residues 1 to 236) in Sf21 cells, and the EDS1-SAG101 complex was similarly purified. Purified GST-NRG1A protein expressed in insect cells was used to pull down the apo-EDS1-SAG101 protein or the EDS1-SAG101 protein coexpressed with RPP1 and ATR1. Small molecules extracted from the coexpressed EDS1-SAG101 protein were incubated with purified apo-EDS1-SAG101 and GST-NRG1A proteins for pull-down analysis.

For semi–in vitro ADPr-ATP–induced EDS1-SAG101-NRG1 association analysis, EDS1-FLAG, SAG101-FLAG, and $NRG1A^{E14A/E27A}$-SH were coexpressed in *Nb-epss* lacking *EDS1* and its paralogs. Total protein from the plant materials was extracted and incubated with ADPr-ATP. Co-immunoprecipitation (Co-IP) analyses were conducted to detect EDS1-SAG101 interaction with NRG1A in the presence or absence of ADPr-ATP.

EDS1-SAG101 coexpressed with the RPP1 resistosome was purified and denatured to extract the EDS1-SAG1010–bound small molecules for analysis with an orbitrap mass spectrometer coupled with a Vanquish ultrahigh-performance liquid chromatography (UHPLC) system. A bridged ethyl hybrid (BEH) amide column was used for LC separation. Data-dependent MS/MS acquisition (DDA) with full scan of m/z 200 to 1500 in negative ion mode was performed. For detection of TIR-catalyzed ADPr-ATP in *Nb* tobacco, wild-type *RBA1* ($RBA1^{WT}$) and $RBA1^{E86A}$ were transiently expressed in *Nb* tobacco *epss* leaves. *Nb* tobacco *epss* leaf powders of $RBA1^{WT}$ or $RBA1^{E86A}$ were mixed with insect cells expressing EDS1-SAG101 for sonication. EDS1-SAG101 was purified, the purified EDS1-SAG101 complex protein was denatured, and the extractant was used for LC-MS analysis with an LC MS-8060 triple quadrupole mass spectrometer.

To detect TIR-catalyzed products with different substrates, purified RPS4-TIR protein bound in Ni-NTA beads was first incubated with 1 ml of ddH_2O containing 1 mM NAD^+, NAD^+ + ADP, or NAD^+ + ATP, and then with purified apo-EDS1-PAD4 or apo-EDS1-SAG101. Then, 200 μl of each reaction solution was incubated with ADR1-L1 or NRG1A for pull-down analysis using Glutathione Sepharose 4B (GS4B) beads. The remaining solution of each

reaction was used for analysis with an orbitrap mass spectrometer.

Cryo-EM data from frozen hydrated grids of the EDS1-SAG101 protein coexpressed with the RPP1 resistosome were collected on a Titan Krios3 (FEI) with K3 direct detection camera (Gatan). Particle picking, two-dimensional (2D) classification, 3D classification, and refinement were all performed in RELION-3.1. The crystal structure of apo-EDS1-SAG101 [Protein Data Bank (PDB) code: 4NFU] was docked into the finally refined 3D reconstruction, and manual atomic model building was carried out in Coot. Phenix.real_space.refine was used to refine the model of EDS1-SAG101 with the final EM map.

N-terminally GST tagged $RPP1^{TIR}$ expressed in sf21 insect cells and purified by GST4B resin was concentrated for crystal screening. The $RPP1^{TIR}$ crystals were obtained, and the crystal structure of $RPP1^{TIR}$ was determined by molecular replacement. The model from MR was built with Coot and subsequently subjected to refinement by Phenix_real_space_refine.

Detailed descriptions of the total synthesis of the di-ADPR enantiomer can be found in data S1. All compounds were confirmed by nuclear magnetic resonance (NMR) spectra (data S1, NMR spectra of compounds).

REFERENCES AND NOTES

1. B. P. M. Ngou, P. Ding, J. D. G. Jones, Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34**, 1447–1478 (2022). doi: 10.1093/plcell/koac041; pmid: 35167697
2. D. Lapin, O. Johanndrees, Z. Wu, X. Li, J. E. Parker, Molecular innovations in plant TIR-based immunity signaling. *Plant Cell* **34**, 1479–1496 (2022). doi: 10.1093/plcell/koac035; pmid: 35143666
3. L. Wan *et al*., TIR domains of plant immune receptors are NAD^+- cleaving enzymes that promote cell death. *Science* **365**, 799–803 (2019). doi: 10.1126/science.aax1771; pmid: 31439793
4. S. Horsefield *et al*., NAD^+ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019). doi: 10.1126/science.aax1911; pmid: 31439792
5. A. M. Bayless, M. T. Nishimura, Enzymatic Functions for Toll/Interleukin-1 Receptor Domain Proteins in the Plant Immune System. *Front. Genet.* **11**, 539 (2020). doi: 10.3389/fgene.2020.00539; pmid: 32582284
6. J. D. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016). doi: 10.1126/science.aaf6395; pmid: 27934708
7. B. P. M. Ngou, H. K. Ahn, P. Ding, J. D. G. Jones, Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* **592**, 110–115 (2021). doi: 10.1038/s41586-021-03315-7; pmid: 33692545
8. M. Yuan *et al*., Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* **592**, 105–109 (2021). doi: 10.1038/s41586-021-03316-6; pmid: 33692546
9. M. T. Nishimura *et al*., TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2053–E2062 (2017). doi: 10.1073/pnas.1620973114; pmid: 28137883
10. M. E. Santamaría *et al*., An Arabidopsis TIR-Lectin Two-Domain Protein Confers Defense Properties against *Tetranychus urticae*. *Plant Physiol.* **179**, 1298–1314 (2019). doi: 10.1104/pp.18.00951; pmid: 30765478
11. T. Zhao *et al*., A truncated NLR protein, TIR-NBS2, is required for activated defense responses in the *exo70B1* mutant. *PLOS Genet.* **11**, e1004945 (2015). doi: 10.1371/journal.pgen.1004945; pmid: 25617755
12. S. Ma *et al*., Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**, eabe3069 (2020). doi: 10.1126/science.abe3069; pmid: 33273071
13. R. Martin *et al*., Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020). doi: 10.1126/science.abd9993; pmid: 33273074
14. Z. Duxbury *et al*., Induced proximity of a TIR signaling domain on a plant-mammalian NLR chimera activates defense in plants. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18832–18839 (2020). doi: 10.1073/pnas.2001185117; pmid: 32709746
15. S. Wagner *et al*., Structural basis for signaling by exclusive EDS1 heteromeric complexes with SAG101 or PAD4 in plant innate immunity. *Cell Host Microbe* **14**, 619–630 (2013). doi: 10.1016/j.chom.2013.11.006; pmid: 24331460
16. D. Lapin *et al*., A coevolved EDS1-SAG101-NRG1 module mediates cell death signaling by TIR-domain immune receptors. *Plant Cell* **31**, 2430–2455 (2019). doi: 10.1105/tpc.19.00118; pmid: 31311833
17. J. Gantner, J. Ordon, C. Kretschmer, R. Guerois, J. Stuttmann, An EDS1-SAG101 complex is essential for TNL-mediated immunity in *Nicotiana benthamiana*. *Plant Cell* **31**, 2456–2474 (2019). doi: 10.1105/tpc.19.00099; pmid: 31266900
18. V. Bonardi *et al*., Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16463–16468 (2011). doi: 10.1073/pnas.1113726108; pmid: 21911370
19. O. X. Dong *et al*., TNL-mediated immunity in *Arabidopsis* requires complex regulation of the redundant *ADR1* gene family. *New Phytol.* **210**, 960–973 (2016). doi: 10.1111/nph.13821; pmid: 27074399
20. T. Qi *et al*., NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10979–E10987 (2018). doi: 10.1073/pnas.1814856115; pmid: 30373842
21. B. Castel *et al*., Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* **222**, 966–980 (2019). doi: 10.1111/nph.15659; pmid: 30582759
22. S. C. Saile *et al*., Two unequally redundant "helper" immune receptor families mediate *Arabidopsis thaliana* intracellular "sensor" immune receptor functions. *PLOS Biol.* **18**, e3000783 (2020). doi: 10.1371/journal.pbio.3000783; pmid: 32925907
23. X. Sun *et al*., Pathogen effector recognition-dependent association of NRG1 with EDS1 and SAG101 in TNL receptor immunity. *Nat. Commun.* **12**, 3335 (2021). doi: 10.1038/s41467-021-23614-x; pmid: 34099661
24. H. Cui *et al*., A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in *Arabidopsis* immunity. *New Phytol.* **213**, 1802–1817 (2017). doi: 10.1111/nph.14302; pmid: 27861989
25. D. D. Bhandari *et al*., An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in *Arabidopsis*. *Nat. Commun.* **10**, 772 (2019). doi: 10.1038/s41467-019-08783-0; pmid: 30770836
26. Z. Wu, L. Tian, X. Liu, Y. Zhang, X. Li, TIR signal promotes interactions between lipase-like proteins and ADR1-L1 receptor and ADR1-L1 oligomerization. *Plant Physiol.* **187**, 681–686 (2021). doi: 10.1093/plphys/kiab305; pmid: 34608964
27. J. A. Dongus, J. E. Parker, EDS1 signalling: At the nexus of intracellular and surface receptor immunity. *Curr. Opin. Plant Biol.* **62**, 102039 (2021). doi: 10.1016/j.pbi.2021.102039; pmid: 33930849
28. S. Huang *et al*., Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity. *Science* **377**, eabq3297 (2022). doi: 10.1126/science.abq3297
29. A. P. Rehmany *et al*., Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1839–1850 (2005). doi: 10.1105/tpc.105.031807; pmid: 15894715
30. P. Jacob *et al*., Plant "helper" immune receptors are Ca^{2+}-permeable nonselective cation channels. *Science* **373**, 420–425 (2021). doi: 10.1126/science.abg7917; pmid: 34140391
31. L. M. Jubic, S. Saile, O. J. Furzer, F. El Kasmi, J. L. Dangl, Help wanted: Helper NLRs and plant immune responses. *Curr. Opin. Plant Biol.* **50**, 82–94 (2019). doi: 10.1016/j.pbi.2019.03.013; pmid: 31063902
32. J. A. Dongus *et al*., Cavity surface residues of PAD4 and SAG101 contribute to EDS1 dimer signaling specificity in plant immunity. *Plant J.* **110**, 1415–1432 (2022). doi: 10.1111/tpj.15747; pmid: 35324052
33. M. R. Swiderski, D. Birker, J. D. Jones, The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. *Mol. Plant Microbe Interact.* **22**, 157–165 (2009). doi: 10.1094/MPMI-22-2-0157; pmid: 19132868
34. Y. Shi *et al*., Structural basis of SARM1 activation, substrate recognition, and inhibition by small molecules. *Mol. Cell* **82**, 1643–1659.e10 (2022). doi: 10.1016/j.molcel.2022.03.007; pmid: 35334231
35. D. Yu *et al*., TIR domains of plant immune receptors are 2′,3′-cAMP/cGMP synthetases mediating cell death. *Cell* **185**, 2370–2386.e18 (2022). doi: 10.1016/j.cell.2022.04.032; pmid: 35597242
36. E. E. Alemasova, O. I. Lavrik, Poly(ADP-ribosyl)ation by PARP1: Reaction mechanism and regulatory proteins. *Nucleic Acids Res.* **47**, 3811–3827 (2019). doi: 10.1093/nar/gkz120; pmid: 30799503
37. Y. He, H. Hara, G. Núñez, Mechanism and Regulation of NLRP3 Inflammasome Activation. *Trends Biochem. Sci.* **41**, 1012–1021 (2016). doi: 10.1016/j.tibs.2016.09.002; pmid: 27669650
38. B. P. M. Ngou *et al*., Estradiol-inducible AvrRps4 expression reveals distinct properties of TIR-NLR-mediated effector-triggered immunity. *J. Exp. Bot.* **71**, 2186–2197 (2020). doi: 10.1093/jxb/erz571; pmid: 32050020
39. R. S. Nandety *et al*., The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* **162**, 1459–1472 (2013). doi: 10.1104/pp.113.219162; pmid: 23735504
40. H. Tian *et al*., Activation of TIR signalling boosts pattern-triggered immunity. *Nature* **598**, 500–503 (2021). doi: 10.1038/s41586-021-03987-1; pmid: 34544113
41. M. Chodasiewicz *et al*., 2′,3′-cAMP treatment mimics the stress molecular response in *Arabidopsis thaliana*. *Plant Physiol.* **188**, 1966–1978 (2022). doi: 10.1093/plphys/kiac013; pmid: 35043968

ACKNOWLEDGMENTS

We thank the National Protein Science Facility at Tsinghua University for technical assistance; J. Lei, X. Li, X. Fan, and N. Liu at Tsinghua University for data collection; and the Tsinghua University Branch of the China National Center for Protein Sciences (Beijing) for providing the cryo-EM facility support and the computational facility support on the cluster of Bio-Computing Platform. We thank O. Johanndrees at MPI-Cologne for providing the EDS1, PAD4, and SAG101 ortholog sequences. We are grateful to G. A. Leonard at the ESRF for providing assistance with beamline ID23-2. **Funding:** This work was supported by the National Key Research and Development Program of China 2021YFA1300701 (Z.H.), the Alexander von Humboldt Foundation (a Humboldt professorship to J.Chai), the Max-Planck-Gesellschaft (J.E.P. and a Max Planck fellowship to J.Chai), Deutsche Forschungsgemeinschaft SFB-1403-414786233 (J.Chai and J.E.P.) and Germany's Excellence Strategy CEPLAS (EXC-2048/1, project 390686111) (J.Chai and J.E.P.), an Alexander von Humboldt Foundation postdoctoral fellowship (J.W.), International Max-Planck Research School (IMPRS) PhD fellowships (H.L. and G.H.), and the National Natural Science Foundation of China (82130103 and U1804283) (J.Chang). Crystals were grown in the Cologne Crystallization facility, supported by German Research Foundation grant no. INST 216/682-1 FUGG. **Author contributions:** Experimental design: J.Chai, J.E.P., J.Chang, Z.H., A.J., and S.H. Recombinant protein expression assays, purification, structure determinations, modeling, and data analysis: A.J., S.H., Y.S., W.S., Z.H., and J.Chai. Biochemical assays: A.J., S.H., W.S., J.W., H.L., G.H., J.J., L.X., X.L., E.L., S.M., D.Y., J.G., U.B., Z.H., J.E.P., and J.Chai. Plant expression and Western blot analysis: J.W., H.L., G.H., and J.E.P. Synthesis of small molecules: Y.M., J.H., T.Z., W.Y., and J.Chang. Data analysis: all authors. Manuscript writing: drafted by J.Chai and edited by J.E.P., J.Chang, Z.H., A.J., and S.H. with contributions from other authors. **Competing interests:** A patent application (CN 202210087257.X) has been filed relating to this work. The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text, supplementary materials, or the listed PDB files. For ADPr-ATP–bound EDS1-SAG101 and the crystal structure of $RPP1^{TIR}$, the atomic coordinates have been deposited in the PDB with accession codes 7XJP and 7XOZ, respectively. The EM map for ADPr-ATP–bound EDS1-SAG101 has been deposited in the Electron Microscopy Database with accession code EMD-3323. Plasmids for in planta expression are available from J.E.P. under a material transfer agreement with the Max-Planck Institute for Plant Breeding Research. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq8180
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (*42*–*56*)
MDAR Reproducibility Checklist
Data S1

Submitted 3 May 2022; accepted 22 June 2022
Published online 7 July 2022
10.1126/science.abq8180

RESEARCH ARTICLES

BIOPHYSICS

Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics

Veer I. P. Keizer[1,2,3]†, Simon Grosse-Holz[1,2,4], Maxime Woringer[1,2], Laura Zambon[1,2,3]‡, Koceila Aizel[2]§, Maud Bongaerts[2], Fanny Delille[5], Lorena Kolar-Znika[1,2], Vittore F. Scolari[1,2], Sebastian Hoffmann[3]¶, Edward J. Banigan[4], Leonid A. Mirny[1,4], Maxime Dahan[2]#, Daniele Fachinetti[3]*, Antoine Coulon[1,2]*

Our understanding of the physical principles organizing the genome in the nucleus is limited by the lack of tools to directly exert and measure forces on interphase chromosomes in vivo and probe their material nature. Here, we introduce an approach to actively manipulate a genomic locus using controlled magnetic forces inside the nucleus of a living human cell. We observed viscoelastic displacements over micrometers within minutes in response to near-piconewton forces, which are consistent with a Rouse polymer model. Our results highlight the fluidity of chromatin, with a moderate contribution of the surrounding material, revealing minor roles for cross-links and topological effects and challenging the view that interphase chromatin is a gel-like material. Our technology opens avenues for future research in areas from chromosome mechanics to genome functions.

Recent progress in observing and perturbing chromosome conformation has led to an unprecedented understanding of the physical principles at play in shaping the genome in four dimensions (4D) (*1*). From genomic loops and topologically associating domains to spatially segregated A/B compartments and chromosome territories, the different levels of organization of the eukaryotic genome are thought to arise from various physical phenomena, including phase separation (*2*–*4*), ATP-dependent motors (*4*, *5*), and polymer topological effects (*6*). Nonetheless, the physical nature of chromatin and chromosomes inside the nucleus and its functional implications for mechanotransduction remain an active area of investigation (*7*, *8*). Observation-based studies assessing the mobility of the genome in living cells, from single loci (*9*–*11*) and small regions (*12*) to large domains (*13*), underline the possible existence of different material states of chromatin (e.g., liquid, solid, and gel-like). Extranuclear mechanical perturbations, including whole-nucleus stretching (*14*, *15*), micropipette aspiration (*16*), and application of local pressures (*16*, *17*) or torques (*18*) onto a cell, all affect the overall geometry of the nucleus and reveal global viscoelastic properties. Conversely, intranuclear mechanical manipulation of the genome is rare and technically challenging (*8*). Viscoelasticity measurements using a microinjected 1-μm bead suggested that interphase chromatin may be a cross-linked polymer network (i.e., a gel) (*19*). Recently, intranuclear mechanics were elegantly probed by monitoring the fusion of both artificial (*20*) and naturally occurring (*21*) droplet-like structures. Active mechanical manipulation of an intranuclear structure was recently achieved using an optical tweezer to displace a whole nucleolus in oocytes (*22*) and using optically induced thermophoretic flows within prophase (*23*) or interphase (*24*) nuclei. However, these approaches are limited to the manipulation of large structures or do not apply forces directly on chromatin. These limitations have made it difficult to disentangle various effects (e.g., mechanical response of the nucleus versus chromatin itself or hydrodynamics versus polymer viscoelasticity), leading to contradictory results. Therefore, an approach for the direct and active mechanical manipulation of specific genomic loci inside living cells is needed. To meet this need, we developed a technique for targeted micromanipulation of a specific genomic locus in the nucleus of a living cell, allowing us to probe the physical properties of an interphase chromosome by measuring its response to a controlled point force.

[1]Institut Curie, PSL Research University, Sorbonne Université, CNRS UMR3664, Laboratoire Dynamique du Noyau, 75005 Paris, France. [2]Institut Curie, PSL Research University, Sorbonne Université, CNRS UMR168, Laboratoire Physico Chimie Curie, 75005 Paris, France. [3]Institut Curie, PSL Research University, CNRS UMR144, Laboratoire Biologie Cellulaire et Cancer, 75005 Paris, France. [4]Department of Physics and Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]ESPCI Paris, PSL Research University, Sorbonne Université, CNRS UMR8213, Laboratoire de Physique et d'Étude des Matériaux, 75005 Paris, France.
***Corresponding author. Email: antoine.coulon@curie.fr (A.C.); daniele.fachinetti@curie.fr (D.F.)**
†Present address: National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.
‡Present address: Lifeanalytics srl, Oderzo, Veneto, Italy.
§Present address: TreeFrog Therapeutics, Pessac, France.
¶Present address: Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany.
#Deceased.

Results

Mechanical manipulation of a genomic locus in a living cell

Our approach relies on tethering magnetic nanoparticles (MNPs) to a genomic locus and applying an external magnetic field (Fig. 1A). We chose ferritin MNPs for their small size (*25*, *26*): 12 nm in diameter for ferritin (PDB 1GWG) and 28 nm for the full MNP (*26*). We produced ferritin MNPs by synthesizing in vitro recombinant enhanced green fluorescent protein (eGFP)–labeled ferritin cages and loading them with a magnetic core (see the materials and methods). We microinjected MNPs into the nuclei of living human U-2 OS cells previously engineered to contain an artificial array inserted at a single genomic location in a subtelomeric region of chromosome 1 (band 1p36) (*27*). This genomic array contains ~200 copies of a 20-kb genetic construct, each including 96 *tetO* binding sites and a transgene. It has been extensively used in the past to study the function of several chromatin modifications, RNA polymerase II (Pol II) recruitment, and RNA synthesis during induction of the transgene (*27*–*29*). Therefore, although we used it here uninduced, this array can recapitulate functional chromatin-based processes such as transcriptional activation. MNPs were targeted to the array using a constitutively expressed fusion protein (TetR, mCherry, and anti-GFP nanobody) serving as a tether (Fig. 1A). Upon injection, MNPs diffused through the nucleus and accumulated at the array, forming a fluorescent spot in both eGFP and mCherry channels (figs. S1A and S2A). Quantification of the fluorescence signals indicated that MNPs were at nanomolar concentrations in the nucleus after injection and accumulated at the genomic locus in the range of hundreds to thousands of MNPs (median 1500 MNPs; figs. S1B and S3, movie S1, table S1, and materials and methods). The locus should be regarded as a condensed and heterochromatic 4-Mb region (1.6% of chromosome 1) residing in a euchromatic genomic context, as previously reported (*27*), with small MNPs (each being ~2 to 3 times (*26*) the size of a nucleosome) sparsely decorating chromatin (1 MNP per ~2.7 kb). Consistently, we observed that the locus typically resided in low to intermediate DNA density regions and was itself relatively condensed (fig. S4, A and B) and that binding of MNPs to the locus did not substantially affect its morphology (fig. S2, A and B). Microinjection and attraction of unbound MNPs did not substantially alter chromatin distribution and densities inside the nucleus [as assessed by silicon rhodamine (SiR)–DNA staining; fig. S2, C and D]. Cells were imaged on a coverglass with custom-made microfabricated pillars (*30*, *31*) that behave as local magnets only when subjected to an external magnetizing field (fig. S5). Therefore, ON/OFF

modulation of the local force field could be achieved while imaging by placing and removing an external magnet on the microscope stage. The shape and orientation of the pillars were chosen to maximize the magnetic field gradient and hence the force. We performed magnetic simulations and experimental calibrations using two independent methods (see the materials and methods, figs. S6 and S7, and movie S2) to determine the magnitude and orientation of the force applied onto the genomic locus as a function of the number of MNPs bound to it and its position relative to the magnetic pillar (Fig. 1B). The typical forces applied onto the locus were in the subpiconewton (pN) range, occasionally reaching a few piconewtons (table S1; median force = 0.45 pN). These values are in the range of

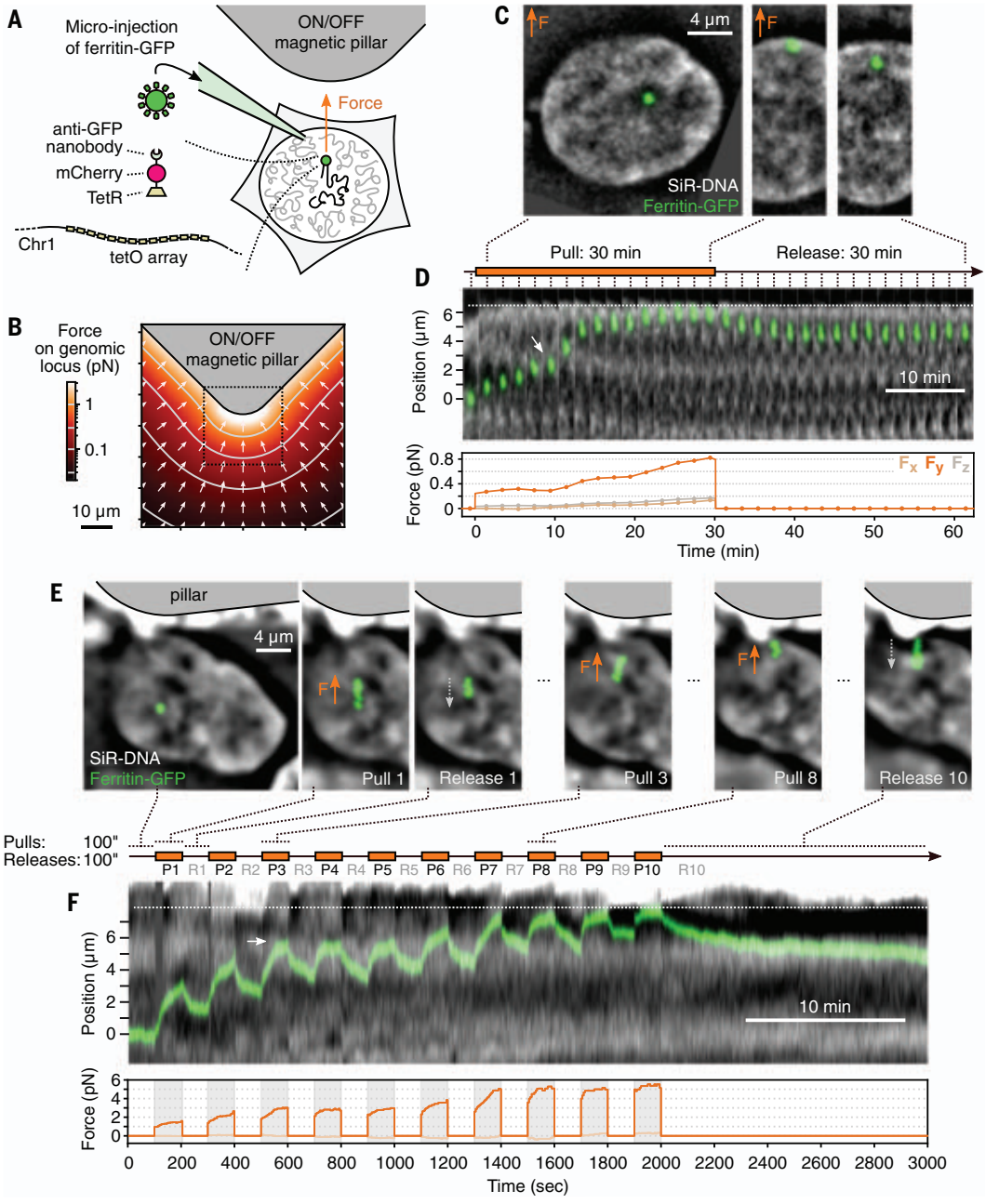


Fig. 1. Mechanical micromanipulation of a genomic locus in living cells. (**A**) MNPs of GFP-labeled ferritin were microinjected into the cell nucleus and targeted to a genomic array containing ~19,000 *tetO* binding sites (*27*) with a linker protein. Cells were imaged on a coverslide with microfabricated magnetic pillars that produce a local magnetic field and attract the genomic locus. (**B**) The force exerted onto the locus depends on its position relative to the pillar and is characterized using a precalculated force map (see the materials and methods), here shown for 1000 MNPs at the locus. (**C**) Example of a pull-release experiment showing the locus being displaced during the 30 min of force exertion and recoiling during the 30 min of force release (30′-PR scheme). (See also movie S3.) (**D**) Kymograph of the same experiment showing each time frame, along with the force time profile calculated using the force map. (**E**) Experiment in which pulls and releases are 100 s and the pull-release cycle is repeated 10× (100″-PR scheme). Images are time projections; shown in green are all positions of the center of mass of the locus over the periods represented on the timeline. The arrows indicate the direction of the motion. (See also movie S4.) (**F**) Kymograph of the same experiment showing the displacement of the locus and the spatial patterns of DNA density in the nucleus, along with the force time profile. All SiR-DNA images are band-pass filtered (see the materials and methods). For (D) and (F), dotted lines indicate the nuclear periphery and white arrows features of interest in the spatial distribution of DNA density.

forces exerted by molecular motors in the nucleus, e.g., comparable to the stalling force of ~0.5 pN for the structural maintenance of chromosomes (SMC) complex condensin (*32*) and a few piconewtons for Pol II (*33*).

Force-induced movement of a genomic locus reveals viscoelastic properties of chromatin

We first applied the magnetic force for 30 min and released it for another 30 min while performing low-illumination three-dimensional imaging with a 2-min interval (30′-PR scheme). We observed a clear motion of the locus toward the magnet upon application of the force and a slow and partial recoil when the force stopped (Fig. 1, C and D, and movie S3). This indicates that a sub-piconewton force, when applied in a sustained and unidirectional manner on a genomic locus, elicits a displacement of that locus by several micrometers over minutes. It also shows that the chromosomal locus can move across the nuclear environment, which is believed to be crowded and entangled. We also applied the force periodically, pulling for 100 s, releasing for 100 s, and repeating this cycle 10 times (100″-PR scheme) while performing fast two-dimensional imaging with a 5-s interval (Fig. 1, E and F, and movie S4). Several observations from these two experiments hinted at the material properties of chromatin. First, the trajectories showed recoils during release periods and a gradual slowdown during both pulls and releases, characteristic of a viscoelastic material. Second, spatial heterogeneities in the trajectories were visible and appeared to relate to the spatial distribution of DNA density (the motion of the locus was occasionally hindered where the DNA density varied; Fig. 1, D and F, white arrows). Third, recoil after force release was seen even after collision with the nuclear periphery, indicating that the material there (peripheral heterochromatin and the nuclear lamina) was not sticky enough to fully retain the locus. Fourth, the spatial distribution of DNA density in the nucleus did not show large-scale deformations, indicating that the locus did not drag along large amounts of material (movies S3 and S4). The force-induced displacements that we observed are consistent with viscoelastic and nonconfining chromatin and constitute a basis to further develop and test physical models of interphase chromosomes.

Quantitative force-response and scaling laws of interphase chromatin mechanics

To quantify the viscoelastic properties of chromatin, we analyzed the trajectories of the locus in 35 cells undergoing the 30′-PR scheme (corrected for cell motion and force orientation; see the materials and methods). We observed a range of behaviors in both pulls and releases regarding initial speed, total distance traveled, and shape of time profiles (Fig. 2, A and B, and fig. S8). Most traces showed a displacement that was clearly distinguishable from diffusion (Fig. 2, A and B, hatched areas; see the materials and methods). Collision with the nuclear periphery (Fig. 2, A and B, open symbols, and fig. S8, open symbols) was seen in nine of 35 traces, so the total displacement during the pull is most often not limited by the nuclear periphery. The initial force applied onto the locus largely predicted the variability seen in the initial motion (Fig. 2C and fig. S9A). The recoil motion after force release was in part predicted by the total distance over which the locus had been displaced during the pull (Fig. 2D and fig. S9B), with a simple linear relationship highlighting the elastic nature of chromatin. Deviations from these simple proportionality relationships indicate that the specific nuclear context or the state of the genomic locus might influence its response. In particular, we observed that when the locus moved more slowly than expected, it was less DNA dense, and when the locus moved faster than expected, it was more DNA dense (fig. S4C), suggesting that the compaction state of the locus itself affected its response to the force. The absolute nuclear position of the locus did not correlate with its response to the force, but if the locus reached the periphery during the pull, it often recoiled more slowly than expected (fig. S10). Despite the variability between traces, log-log plots of all the pulls and releases from the 30′-PR and 100″-PR trajectories, together with three additional high-frame-rate (dt = 0.5″) pull-release trajectories, revealed linear portions in the curves with a slope of 0.5 over more than three orders of magnitude in time (Fig. 2E). This behavior suggests that the different levels of the hierarchical genome organization are not characterized by vastly distinct mechanical properties. In addition, displacements that scale with time as $t^{1/2}$ can be empirically described by a "fractional speed," i.e., a single value in $\mu m/s^{1/2}$ capturing how the motion evolves over time (Fig. 2, C and D, and fig. S9, right axes). The first pull of the 100″-PR trace, represented in this unit, indeed followed the same relationship as the 30′-PR traces (Fig. 2C, dark green triangle), and the slope of the resulting force-displacement plot yielded a unique factor of 0.158 (±0.014) $\mu m/s^{1/2}/pN$, characterizing the dynamic response of chromatin to force. These results indicate that a large part of the response of chromatin to force can be described by simple laws.

The chromatin force response is well described by a free polymer model (Rouse chain)

We then sought a model of chromatin that best explains our quantitative measurement of force-induced locus displacement. Several features in our data suggested a classical polymer model known as a Rouse polymer (*34*) as a first approximation to describe the response of chromatin to forces. The Rouse model represents a polymer in which each monomer diffuses by thermal motion in a viscous medium and is connected to its two neighbors by elastic bonds. Rouse ignores steric effects (contact, hindrance), cross-links (affinity, stickiness), and topological effects (fibers can pass through each other). This model is frequently invoked for chromatin dynamics because it predicts the characteristic power-law scaling–that is, a linear relationship on a log-log plot–of the mean squared displacement (MSD) versus time with exponent 0.5, as observed here (fig. S11, A and E) and for other genomic loci in eukaryotes (*35*–*37*). We extended Rouse theory to study how a polymer responds to a point force (see the materials and methods and the supplementary text). Our calculations predict a power-law behavior with exponent 0.5 for displacements and recoils in response to force, consistent with our experimental observations (Fig. 2E). These two power laws have the same physical origin, so the diffusion coefficient obtained independently from the MSD ($1627 \pm 19\ nm^2\ s^{-1/2}$; fig. S11A) directly relates to—and predicts—the slope of the force-displacement plot (Fig. 2C, red line): $1627\ nm^2\ s^{-1/2}/2k_BT = 0.190 \pm 0.003\ \mu m/s^{1/2}/pN$ (see the materials and methods). This agreement between two independent passive and active measurements (Fig. 2C and fig. S9A, red and gray lines, representing diffusion and force response, respectively) supports the Rouse model to explain our chromatin dynamics data. Inspected on a cell-by-cell basis, the force-free MSD of the locus before and after the pull-release experiments appeared very moderately reduced in most cases (fig. S11B). Its natural variability between cells does not appear to explain the variability of the response to force (fig. S11C). After force release, the Rouse model also predicted a recoil proportional to the total displacement during the pull. However, in many cases, the locus recoiled somewhat more slowly than predicted by the Rouse theory. Instead, the theoretical prediction appears to define an upper bound for the recoils (Fig. 2D and fig. S9B, red lines), and deviations from Rouse theory were more pronounced at the nuclear periphery (fig. S10B). This analysis suggests that the dynamic response of the chromosome to the force can be described by the Rouse polymer model, with additional effects from the nuclear environment.

Model-based trajectory analysis reveals moderate hindrance by surrounding chromatin

To further understand the physical nature of chromatin, we investigated how alternative polymer models are able to capture the 100″-PR trace (Fig. 3, A to D). Our approach was to use the displacement trajectory and infer, assuming a given polymer model, the time profile of

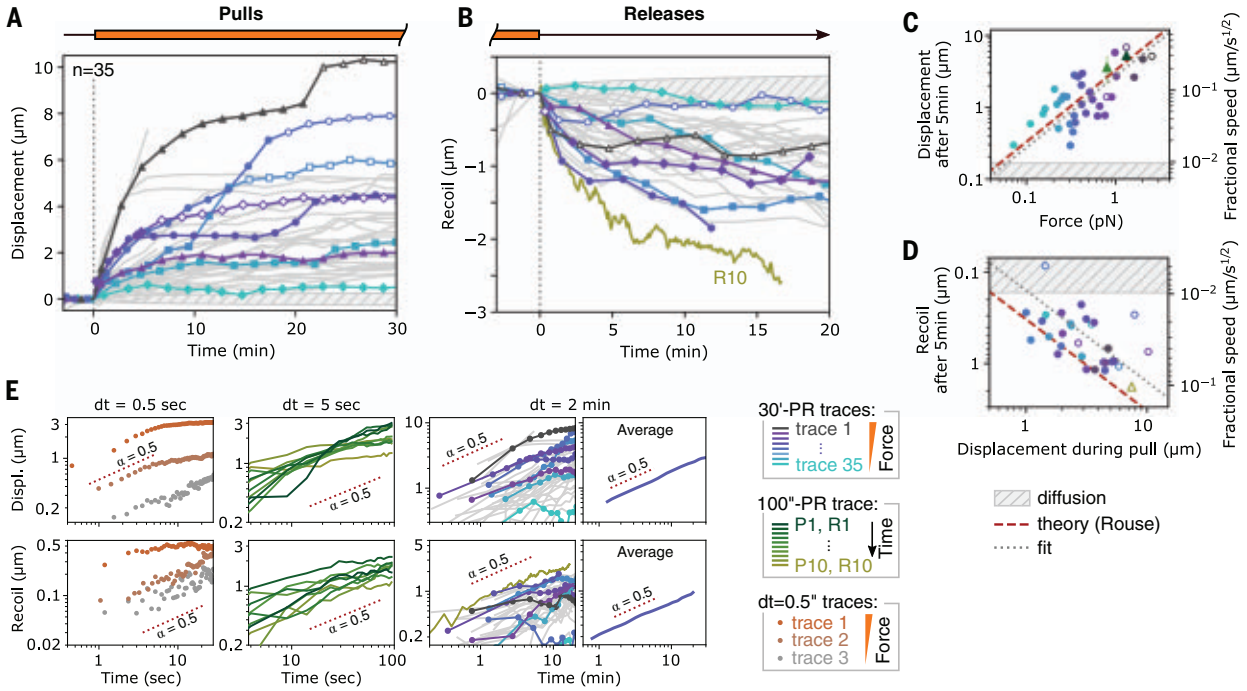


Fig. 2. Quantitative analysis of locus movement in response to force. (**A**) Trajectories of the genomic locus in the direction of the applied force for 35 different cells during force exertion with the 30′-PR scheme. A selection of trajectories, representative of the breadth of observed behaviors, are highlighted and color-coded by force (see fig. S8 for individual trajectories). Hatched areas in (A) to (D) correspond to the null model of pure diffusion based on MSD measurement (see the materials and methods). (**B**) Recoil trajectories relative to the time and position at the start of the release are shown for the same loci as in (A). Curve R10 is the last release of the 100″-PR trajectory. (**C**) Displacements measured at Δt = 5 min of force exertion on all the traces from (A) plotted against the magnitude of the force. Coordinates are interpolated between the frames before/after Δt. The green line and triangle correspond to the envelope of the 100″-PR trajectory. Reported forces are the average over Δt. Displacements are also expressed in $\mu m/s^{1/2}$ (right axis), allowing us to place pull P1 from the 100″-PR trace, measured at Δt = 100 s (dark green triangle). The red line indicates the expected relationship from Rouse theory, solely based on an MSD measurement. (**D**) Recoil after Δt = 5 min of force release on all the traces from (B) and the last release of the 100″-PR trajectory (R10, green triangle), plotted against the total displacement during the pull. The red line indicates the expected relationship from Rouse theory [see fig. S9 for versions of (C) and (D) at different Δt and in linear scales]. Open symbols in (A) to (D) indicate when loci are within 1.5 μm of the nuclear periphery [at the moment of measurement in (A) to (C) and at the moment of force release in (D)]. (**E**) Displacement and recoil trajectories, aligned on the time and position at the moment of force switching, are represented as log-log plots for pull-release experiments imaged with different frame intervals: dt = 0.5 s (see the materials and methods), dt = 5 s (from the 100″-PR trace; Fig. 1F), and dt = 2 min [all 30′-PR traces; (A) and (B)]. For the latter, average trajectories (right plots) were calculated over all displacements where the applied force remains ≤2 pN (28 of 35 traces) and over all the recoils after a displacement of ≤5 μm (21 of 35 traces). Red dotted lines indicate the power-law behavior, with exponent 0.5, predicted by Rouse theory.

the force that produced the measured trajectory (see the materials and methods, supplementary text, and fig. S12, A to C). Disagreement between inferred and actual force profiles indicates when models are incorrect or incomplete, allowing one to select and refine the best model(s). With this approach, we compared a series of models (fig. S12C). First, a simple Rouse model without any adjustable parameters (i.e., calibrated using the MSD versus time plot; fig. S11A) predicted well the first pull and all of the release periods (Fig. 3B). However, the prediction leaves some of the applied force unexplained (Fig. 3B, gray area between curves), suggesting a missing component in the model that would additionally slow down or hinder the progression of the locus. This residual unexplained force did not scale with speed and thus could not be explained as an additional viscous drag on the locus (fig. S12D). Instead, it increased progressively across successive pulls, suggesting an accumulation of hindrance as the locus moved through the nucleus. To represent this, we added a capacity for the locus to interact with the surrounding chromatin, represented as extra Rouse chains that are either attached to or pushed by the locus along its path (Fig. 3C and fig. S12C). These models better predicted the force profile throughout the trajectory compared with a pure Rouse model. The only free parameter used for the inference shown in Fig. 3C is the frequency at which the locus interacts with other polymers, which we found to be very low (fig. S12C), indicating that the interaction with the surrounding chromatin was moderate. This is also consistent with the small but detectable reduction in mobility of the locus before and after pull-release experiments (fig. S11B) and the subtle redistribution of DNA densities around the pulled locus (fig. S4D). These modeling results suggest that, upon force application and release on our genomic locus, chromatin is well described as a Rouse polymer (i.e., a free polymer in a viscous environment), with moderate interactions from the surrounding chromatin, indicating that hindrance, cross-links, and topological effects play a minor role.

Interphase chromatin does not behave as a gel in force-response experiments

Interphase chromatin has been proposed to be a gel-like material (*11*, *12*, *19*). A gel is a highly cross-linked polymer, which means that unlike a linear polymer, in which monomers are linked to two neighbors, extra links between nonadjacent monomers form an interconnected mesh, giving the gel solid-like properties. For chromatin, this could in principle arise from affinity between nucleosomes, as well as loops or bridges formed by proteins, complexes, and condensates and topological entanglement between chromatin fibers. However, in such an interconnected mesh structure, short paths effectively linking the pulled locus

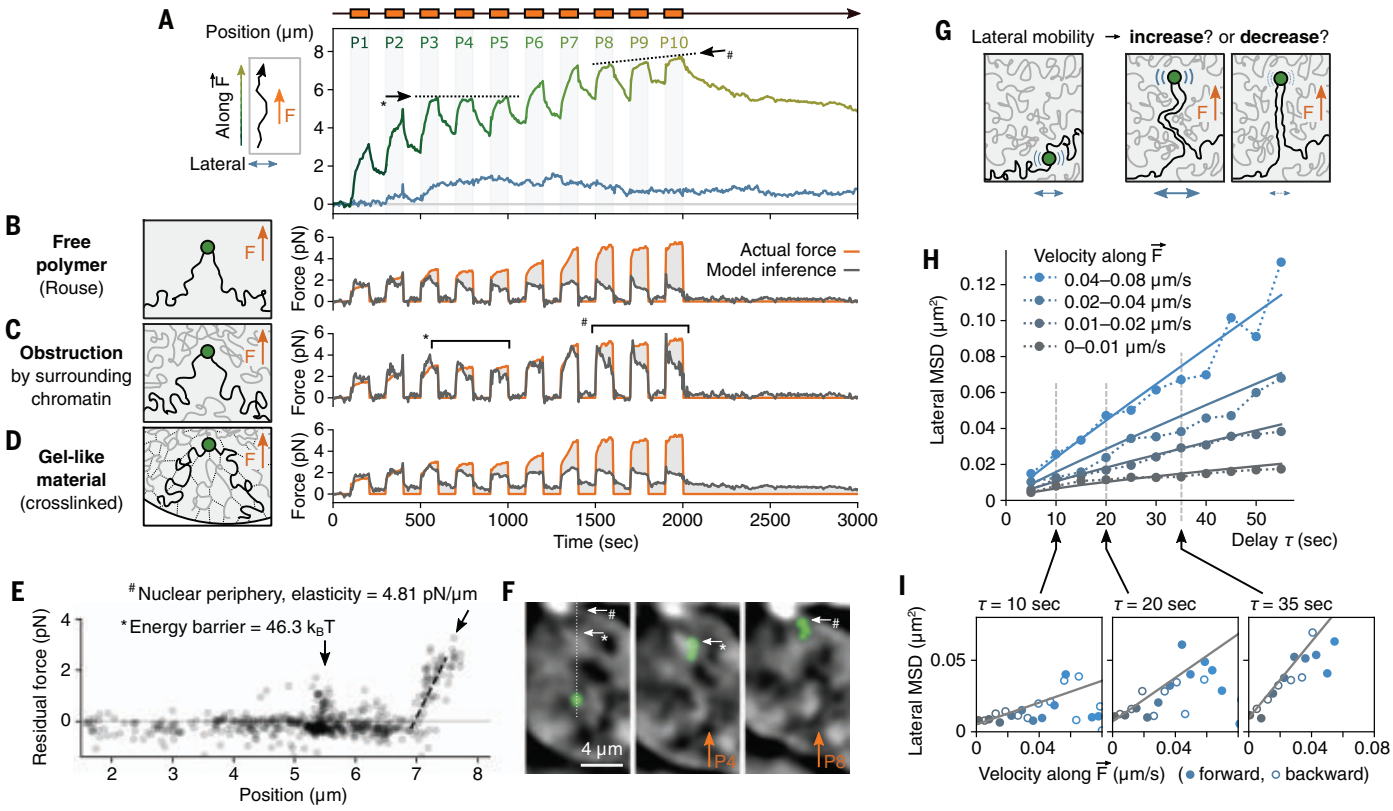


Fig. 3. Model-based analysis and hypothesis testing. (**A**) Trajectory of the locus shown in the direction of the force (green curve) and orthogonal to the force in the imaging plane (blue curve) for the 100″-PR experiment. Arrows indicate apparent obstacle [see asterisks and hash marks in (C), (E), and (F)]. (**B** to **D**) Evaluation of different models through their capacity to reproduce the experimentally measured force time profile (orange curve) by inferring it from the trajectory (gray curve). Models shown here are a simple Rouse polymer (*34*) (B), the same model with extra polymer chains being pushed by the locus to represent the surrounding chromatin (C), and a gel-like material represented as a Rouse polymer in a viscoelastic environment (D). (See full list in fig. S12.) (**E**) Residual unexplained force from the second model [area between curves in (C)] plotted along the trajectory of the locus, highlighting an obstacle (asterisk) and an elastic region near the nuclear periphery (hash mark) for which physical parameters are measured (see the materials and methods) and which are visible in (A), (C), and (F). (**F**) Time projection images before the first pull and during pulls P4 and P8 showing how the spatial distribution of DNA density in the nucleus relates to the identified obstacles. SiR-DNA images were band-pass filtered (see the materials and methods). (**G**) Hypotheses on how the lateral mobility of the locus may change depending on its force-induced displacement. (**H** and **I**) MSD of the lateral movement of the locus calculated as a function of both time delay and velocity in the direction of the force. Solid lines on both the MSD-delay (H) and the MSD-velocity (I) representations correspond to a single-parameter fit describing how lateral mobility increases with velocity in the direction of the force.

to all other loci in the nucleus would result in long-range deformation of the spatial pattern of DNA density, which we did not observe (Fig. 1, D and F, and movies S3 and S4). Further, if the chromatin surrounding the locus were gel-like, then it would effectively act as a viscoelastic medium. This assumption does not recapitulate well the experimental data (even with two free parameters; Fig. 3D and fig. S12C) and is inconsistent with the observed scaling of 0.5 in the MSD (fig. S11, A and E), which argues for a simply viscous and nonelastic medium. Finally, if the locus were part of an interconnected mesh, then short series of links would tether it to large structures (e.g., the periphery and nucleoli). A Rouse model that includes a finite tether does not recapitulate the experimental data (fig. S12C) and is inconsistent with the linear behavior observed in Fig. 2E up to several micrometers. These results suggest again minor effects of cross-links and topological constraints and argue against the view that interphase chromatin behaves like a gel at the spatial and temporal scale of our observations.

Heterogeneities in the trajectory reveal obstacles in the nuclear interior and a soft elastic material at the nuclear periphery

Even the models that best capture the data leave part of the force unexplained (Fig. 3C, gray area). We plotted this residual unexplained force as a function of spatial position (Fig. 3E). This revealed an accumulation of non-null residual forces at specific locations, matching visible features in the trajectory and in the spatial distribution of DNA density in the nucleus. First, the residual force in pulls P3 to P5 corresponds to an apparent obstacle in the trajectory (Fig. 3, A and C, asterisks) occurring at a high-to-low transition of DNA density (Figs. 3F and 1F). It appears as a spatially defined barrier of residual force (Fig. 3E), requiring an energy of ≈ 46 k_BT to overcome. This suggests that, whereas DNA dense regions are not obstacles per se, the interface between high- and low-density regions may constitute a barrier. The energy that we estimated suggests that such barriers may be overcome by ATP-dependent molecular motors (*32*, *33*) but not likely by spontaneous thermal fluctuations. Second, the residual force in pulls P8 to P10 (Fig. 3, A and C, hash marks) corresponds to the collision with structures near the nuclear periphery (Figs. 1F and 3F, hash marks). The observed linear force-distance relationship (Fig. 3E) indicates a solid-like elastic behavior for these structures over at least 600 nm and with a spring constant of 4.81 pN/μm. This is much softer than what was measured by whole-nucleus stretching experiments (*14*, *15*), which could be explained by the small size of the locus and/or the existence of a soft layer of

elastic peripheral components (e.g., heterochromatin and nuclear lamina) rather than the material directly contributing to the structural rigidity of the nucleus.

Lateral mobility of the locus reflects transient collisions with obstacles in the nucleoplasm

To further investigate the material encountered by the locus, we analyzed the lateral motion of the locus as it was being pulled and released (Fig. 3A, blue curve). We hypothesized that collisions with obstacles could increase lateral mobility or, conversely, that the locus being dragged into a more constraining and entangled environment could result in a reduction of its mobility (Fig. 3G). After computing the MSD of the lateral motion as a function of both time delay τ and velocity υ_y along the direction of the force (Fig. 3, H and I, and fig. S11, F and G), we observed a clear increase of lateral mobility when the locus moved (for both forward and backward movements; Fig. 3I), suggesting the existence of obstacles that deflected the motion. This additional mobility in the MSD is captured by a term proportional to υ_y, as would be expected for collisions, and proportional to τ (not $\tau^{0.5}$), as would be expected if the force due to the collision with obstacles persists in the same direction across several frames, indicating the existence of large obstacles. Indeed, in P3 for instance, the lateral motion clearly shows a directional behavior (Figs. 1E and 3A). However, the relationships that we observed (Fig. 3, H and I) held even when excluding all of the time points before P4 (fig. S11, H and I), indicating that the collision with obstacles was widespread throughout the nucleus. These results, together with our observation that very few chromatin fibers appeared to be carried along with the locus, indicate that obstacles are frequently encountered by the locus, but most interactions are weak and transient.

Discussion

Our measurements of how a genomic locus inside the nucleus of a living cell responds to a point force indicates that interphase chromatin has fluid-like properties and behaves as a free polymer. This contrasts with previous studies depicting chromatin as a stiff, cross-linked polymer gel with solid-like properties (*11*, *12*, *19*). Our observation that near-piconewton forces can easily move a genomic locus across the nucleus over a few minutes (Fig. 1, D and F) also contrasts with a previous study reporting confined submicrometer displacements over seconds upon application of 65 to 110 pN forces to a 1-μm bead (*19*). We propose that our results may be reconciled with previous experiments in several ways. First, unlike a micrometer-sized bead, the locus used in our experiments is small and may be deformable enough to pass through the surrounding chromatin. Second, chromatin may contain many small, gel-like patches embedded in a structure with liquid, Rouse-like properties at a larger scale. This is also consistent with our observation that the transiting locus frequently encounters obstacles. Third, chromatin may be a weak gel, i.e., a gel with short-lived cross-links (*11*). Such a gel could continuously maintain a stiff, globally connected network that resists stresses over large length scales while permitting fluid-like motions at smaller scales. Future experiments perturbing chromatin state and chromatin associated proteins will be important to reconcile observed micro- and mesoscale mechanics.

Organization of chromosomes that allows movement of genomic loci across large distances by weak forces could have implications for a range of genome functions. For example, large-scale movements of chromosomes occur during nuclear inversion in rod cell differentiation for nocturnal mammals (*38*). Specific genes undergo long-range directional motion upon transcriptional activation (*39*, *40*). Long and highly transcribed genes can form ~5-μm giant loops, believed to be due to chromosome fiber stiffening (*41*). Certain double-strand break sites undergo large-scale, nuclear F-actin dependent relocation to the nuclear periphery (*42*). These DNA-based biological processes require a nuclear organization in which such movements are possible. Our results reveal the mechanical properties of chromatin in which such large-scale movements would only require weak (near piconewton) forces. Although sustained unidirectional forces are unlikely to occur naturally in the nucleus, the magnitude of the forces and the time scale of force exertion in our experiments are comparable to those of molecular motors such as SMC complexes and Pol II; that is, in the sub-piconewton (*32*) or low-piconewton (*33*) ranges and applied over minutes (e.g., 10 min for Pol II to elongate through a 25-kb gene, 5–30 min for SMC complexes). Therefore, some molecular motors in the nucleus operate in a force range that is sufficient to substantially reorganize the genome in space.

Future work will be important to expand and complement our results. Although the genomic array that we used here is known to be chromatinized and has been used extensively to recapitulate and study functional chromatin-based processes (*27*–*29*), we cannot exclude that its repetitive and artificial nature might prevent some of our measurements from being applicable to nonrepetitive and native regions. Manipulating loci other than a subtelomeric locus on the longest chromosome (chromosome 1) in other genomic contexts (e.g., heterochromatin or euchromatin) and in different cell types will be important to assess the generalizability of our findings in various biological contexts.

Our approach to mechanically manipulate and relocate genomic loci in the nuclear space opens many avenues for future research, from the study of interphase chromosome mechanics to the perturbation of genome functions, including transcription, replication, DNA damage repair, and chromosome segregation. By giving access to physical parameters and revealing fundamental scaling laws to describe chromatin mechanics, our work provides a foundation for future theories of genome organization.

REFERENCES AND NOTES

1. J. Dekker *et al.*, *Nature* **549**, 219–226 (2017).
2. D. Jost, P. Carrivain, G. Cavalli, C. Vaillant, *Nucleic Acids Res.* **42**, 9553–9561 (2014).
3. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, *Cell* **169**, 13–23 (2017).
4. L. A. Mirny, M. Imakaev, N. Abdennur, *Curr. Opin. Cell Biol.* **58**, 142–152 (2019).
5. E. J. Banigan, L. A. Mirny, *Curr. Opin. Cell Biol.* **64**, 124–138 (2020).
6. J. D. Halverson, J. Smrek, K. Kremer, A. Y. Grosberg, *Rep. Prog. Phys.* **77**, 022601 (2014).
7. C. Uhler, G. V. Shivashankar, *Nat. Rev. Mol. Cell Biol.* **18**, 717–727 (2017).
8. A. A. Agbleke *et al.*, *Mol. Cell* **79**, 881–901 (2020).
9. T. Nozaki *et al.*, *Mol. Cell* **67**, 282–293.e7 (2017).
10. B. Chen *et al.*, *Cell* **155**, 1479–1491 (2013).
11. N. Khanna, Y. Zhang, J. S. Lucas, O. K. Dudko, C. Murre, *Nat. Commun.* **10**, 2771 (2019).
12. H. Strickfaden *et al.*, *Cell* **183**, 1772–1784.e13 (2020).
13. A. Zidovska, D. A. Weitz, T. J. Mitchison, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15555–15560 (2013).
14. A. D. Stephens, E. J. Banigan, S. A. Adam, R. D. Goldman, J. F. Marko, *Mol. Biol. Cell* **28**, 1984–1996 (2017).
15. Y. Shimamoto, S. Tamura, H. Masumoto, K. Maeshima, *Mol. Biol. Cell* **28**, 1580–1589 (2017).
16. K. N. Dahl, A. J. Engler, J. D. Pajerowski, D. E. Discher, *Biophys. J.* **89**, 2855–2864 (2005).
17. C. M. Hobson *et al.*, *Mol. Biol. Cell* **31**, 1788–1801 (2020).
18. A. Tajik *et al.*, *Nat. Mater.* **15**, 1287–1296 (2016).
19. A. H. B. de Vries, B. E. Krenn, R. van Driel, V. Subramaniam, J. S. Kanger, *Nano Lett.* **7**, 1424–1427 (2007).
20. Y. Shin *et al.*, *Cell* **175**, 1481–1491.e13 (2018).
21. C. M. Caragine, S. C. Haley, A. Zidovska, *eLife* **8**, e47533 (2019).
22. M. S. Syrchina, A. M. Shakhov, A. V. Aybush, V. A. Nadtochenko, Optical trapping of nucleolus reveals viscoelastic properties of nucleoplasm inside mouse germinal vesicle oocytes. bioRxiv 999342 [Preprint] (2020); https://doi.org/10.1101/2020.03.19.999342.
23. M. Mittasch *et al.*, Active gelation breaks time-reversal-symmetry of mitotic chromosome mechanics. bioRxiv 296566 [Preprint] (2018); https://doi.org/10.1101/296566.
24. B. Seelbinder, M. Jain, E. Erben, S. Klykov, I. D. Stoev, M. Kreysing, Non-invasive chromatin deformation and measurement of differential mechanical properties in the nucleus. bioRxiv 472786 [Preprint] (2021); https://doi.org/10.1101/2021.12.15.472786.
25. C. Monzel, C. Vicario, J. Piehler, M. Coppey, M. Dahan, *Chem. Sci.* **8**, 7330–7338 (2017).
26. D. Liße *et al.*, *Adv. Mater.* **29**, 1700189 (2017).
27. S. M. Janicki *et al.*, *Cell* **116**, 683–698 (2004).
28. T. Tsukamoto *et al.*, *Nat. Cell Biol.* **2**, 871–878 (2000).
29. X. Darzacq *et al.*, *Nat. Struct. Mol. Biol.* **14**, 796–806 (2007).
30. L. Toraille *et al.*, *Nano Lett.* **18**, 7635–7641 (2018).
31. M. Bongaerts *et al.*, *Int. J. Mol. Sci.* **21**, 6560 (2020).
32. M. Ganji *et al.*, *Science* **360**, 102–105 (2018).
33. E. A. Galburt, S. W. Grill, C. Bustamante, *Methods* **48**, 323–332 (2009).
34. P. E. Rouse Jr., *J. Chem. Phys.* **21**, 1272–1280 (1953).
35. B. Gu *et al.*, *Science* **359**, 1050–1055 (2018).
36. J. Li *et al.*, *Cell* **178**, 491–506.e28 (2019).
37. H. Hajjoul *et al.*, *Genome Res.* **23**, 1829–1838 (2013).
38. I. Solovei *et al.*, *Cell* **152**, 584–598 (2013).
39. C.-H. Chuang *et al.*, *Curr. Biol.* **16**, 825–831 (2006).
40. N. Khanna, Y. Hu, A. S. Belmont, *Curr. Biol.* **24**, 1138–1144 (2014).
41. S. Leidescher *et al.*, *Nat. Cell Biol.* **24**, 327–339 (2022).
42. C. P. Caridi *et al.*, *Nature* **559**, 54–60 (2018).

43. V. I. P. Keizer *et al*., Data, software, and documentation for: Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics, Zenodo (2022); https://doi.org/10.5281/zenodo.4626942.

ACKNOWLEDGMENTS

We dedicate this work to the memory of our co-author Maxime Dahan, who died in July 2018. We acknowledge scientific inputs from C. Giovannangeli and J.-P. Concordet (MNHN), T. Pons and N. Lequeux (ESPCI), M. Coppey (Institut Curie), M. Cosentino Lagomarsino (IFOM), and the members of the Coulon and Fachinetti teams. We also acknowledge the current and past members of the Dahan/Coppey/Hajj team and UMR168, including C. Monzel, D. Liβe, J. Manzi, E. Baloul, C. Vicario, and D. Normanno, for their earlier work and technical help on nanoparticles and microinjection, and E. Secret and A. Michel-Tourgis (UMR8234 CNRS) for help with magnetic nanoparticle characterization. We also acknowledge the Flow Cytometry Core Facility of the Institut Curie, the BMBC platform of UMR168, the machine shop of UMR168, the technological platform of the Institut Pierre-Gilles de Gennes (IPGG) for access to microfabrication facilities, the PICT-IBiSA@Pasteur Imaging Facility of the Institut Curie (member of the France Bioimaging National Infrastructure; ANR-10-INBS-04), and the MPBT platform (Physical Measurements at Low Temperatures) of Sorbonne Université. **Funding:** This work received funding from the LabEx CELL(N)SCALE (ANR-11-LABX-0038 and ANR-10-IDEX-0001-02 to M.D., D.F., and A.C.); the Agence Nationale de la Recherche (project CHROMAG, ANR-18-CE12-0023-01 to M.D. and A.C.); the PRESTIGE program of Campus France (PRESTIGE-2018-1-0023 to V.K.); the ATIP-Avenir program of CNRS and INERM, the Plan Cancer of the French ministry for research and health (A.C. and D.F.); the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 757956 to A.C.); the LabEx DEEP (ANR-11-LABX-0044 and ANR-10-IDEX-0001-02 to A.C.); the program Fondation ARC (grant agreement PJA 20161204869 to A.C.); the Institut Curie (D.F. and A.C.); the Centre National de la Recherche Scientifique (CNRS) (A.C. and D.F.); the European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie grant agreement no. 666003 to S.H.); the National Institutes of Health (grant GM114190 to L.A.M.); and the MIT-France Seed Fund (L.A.M. and M.D.). L.A.M. is a recipient of Chaire Blaise Pascal by Île-de-France. Administration. **Author contributions:** M.D., D.F., and A.C. conceptualized the project. M.D., D.F., V.I.P.K., and A.C. obtained funding. V.I.P.K., M.D., D.F., and A.C. designed the experiments. V.I.P.K., L.Z., F.D., and A.C. obtained the data. V.I.P.K., L.Z., K.A., M.B., F.D., L.K.Z., and S.H. produced reagents and provided key technical expertise. M.W., V.I.P.K., and A.C. preformed image analysis. M.W., V.I.P.K., S.G.H., V.F.S., and A.C. analyzed trajectories. S.G.H., E.J.B., and L.A.M. preformed polymer modeling. V.I.P.K., S.G.H., M.W., V.F.S., E.J.B., L.A.M., D.F., and A.C. interpreted the results. D.F. and A.C. supervised the experimental work. A.C. supervised the image analysis work. E.J.B. and L.A.M. supervised the modeling work. A.C. drafted the manuscript with input from V.I.P.K., M.W., D.F., S.G.H., E.J.B., and L.A.M. All authors participated in reviewing and editing the manuscript. **Competing interests:** The authors declare no competing interests. K.A. is currently employed by TreeFrog Therapeutic as a research and development engineer in encapsulation technologies. **Data and materials availability:** Cell lines and plasmids are available upon request. All data and codes are available through GitHub and Zenodo (*43*), as summarized in table S2. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9810
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 and S2
References (*44*–*49*)
Movies S1 to S4
MDAR Reproducibility Checklist

Submitted 13 April 2021; resubmitted 2 May 2022
Accepted 21 June 2022
10.1126/science.abi9810

SOLAR CELLS

Ion-modulated radical doping of spiro-OMeTAD for more efficient and stable perovskite solar cells

Tiankai Zhang[1]†, Feng Wang[1]*†, Hak-Beom Kim[2], In-Woo Choi[2], Chuanfei Wang[3], Eunkyung Cho[4], Rafal Konefal[5], Yuttapoom Puttisong[1], Kosuke Terado[6], Libor Kobera[5], Mengyun Chen[1], Mei Yang[1], Sai Bai[1], Bowen Yang[7,8], Jiajia Suo[7,8], Shih-Chi Yang[9], Xianjie Liu[3], Fan Fu[9], Hiroyuki Yoshida[6,10], Weimin M. Chen[1], Jiri Brus[5], Veaceslav Coropceanu[4], Anders Hagfeldt[7,8], Jean-Luc Brédas[4], Mats Fahlman[3], Dong Suk Kim[2], Zhangjun Hu[1], Feng Gao[1]*

Record power conversion efficiencies (PCEs) of perovskite solar cells (PSCs) have been obtained with the organic hole transporter 2,2′,7,7′-tetrakis(*N*,*N*-di-*p*-methoxyphenyl-amine)9,9′-spirobifluorene (spiro-OMeTAD). Conventional doping of spiro-OMeTAD with hygroscopic lithium salts and volatile 4-*tert*-butylpyridine is a time-consuming process and also leads to poor device stability. We developed a new doping strategy for spiro-OMeTAD that avoids post-oxidation by using stable organic radicals as the dopant and ionic salts as the doping modulator (referred to as ion-modulated radical doping). We achieved PCEs of >25% and much-improved device stability under harsh conditions. The radicals provide hole polarons that instantly increase the conductivity and work function (WF), and ionic salts further modulate the WF by affecting the energetics of the hole polarons. This organic semiconductor doping strategy, which decouples conductivity and WF tunability, could inspire further optimization in other optoelectronic devices.

Metal halide perovskites have achieved impressive power conversion efficiencies (PCEs) in both single-junction (*1*–*6*) and tandem solar cells (*7*, *8*). However, a key challenge limiting their practical application is the trade-off between high efficiency and high stability, an issue determined by not only the perovskite materials (*9*) but also the charge transport layers (*10*). Currently, most high-performance perovskite solar cells (PSCs) with >24% PCE are based on the benchmark hole transport layer (HTL) spiro-OMeTAD doped by lithium bis(trifluoromethane)sulfonimide (LiTFSI) and 4-*tert*-butylpyridine (tBP) (*11*–*13*), a process that limits the stability of these high-efficiency PSCs.

Conventional spiro-OMeTAD doping (Fig. 1A, top) involves LiTFSI to facilitate the generation of spiro-OMeTAD$^{\bullet+}$TFSI$^-$ radicals and tBP to improve dopants' solubility and film morphology (*14*). This recipe usually requires an in situ oxidization process for a period of 10 to 24 hours in air to reach the optimal conductivity and work function (WF) (*15*). However, because of the low doping efficiency of this process, a large amount of dopants and additives (~56 mol % LiTFSI and ~330 mol % tBP) are required to generate ~10 mol % radicals (*16*). The residual LiTFSI, tBP, and by-products (e.g., Li_xO_y) are not only diffusible (*17*) but also sensitive to humidity and heat (*18*, *19*), negatively affecting device stability.

In addition, the complex in situ oxidation process makes it challenging to understand the mechanism of conventional spiro-OMeTAD doping, which limits further development of stable HTLs with high PCEs. Several efforts addressing the spiro-OMeTAD stability issue have focused on the hygroscopic lithium salt. For example, a CO_2 gas-forming treatment was recently used to remove the hygroscopic lithium species (*20*). Less hygroscopic dopants, including metallic salts (*21*–*23*), protic ionic liquids (*24*), and ex situ synthesized spiro-OMeTAD$^{2\bullet+}$(TFSI$^-$)$_2$ radicals (*25*, *26*), have also been used to replace the oxidant LiTFSI. Such doping usually requires the addition of volatile tBP, whose role in the doping process has been controversial (*27*, *28*). In particular, tBP would react with radicals and coordinate with LiTFSI to form by-products (Fig. 1A, top), which not only negatively affects device performance but also prevents a full understanding of the spiro-OMeTAD doping mechanism.

We developed a stable doping strategy for instant and effective doping of spiro-OMeTAD, which we call ion-modulated (IM) radical doping (Fig. 1A, bottom), that relies on two components. One is the presynthesized stable spiro-OMeTAD$^{2\bullet+}$(TFSI$^-$)$_2$ radical (synthesis and characterizations in figs. S1 and S2), which

[1]Department of Physics, Chemistry and Biology (IFM), Linköping University, 58183 Linköping, Sweden. [2]Korea Institute of Energy Research (KIER), Ulsan, Republic of Korea. [3]Laboratory of Organic Electronics, Department of Science and Technology, Linköping University, 60174 Norrköping, Sweden. [4]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA. [5]Institute of Macromolecular Chemistry of the Czech Academy of Sciences, 162 06 Prague 6, Czech Republic. [6]Graduate School of Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan. [7]Laboratory of Photomolecular Science (LSPM), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [8]Department of Chemistry, Ångström Laboratory, Uppsala University, SE-751 20 Uppsala, Sweden. [9]Laboratory for Thin Films and Photovoltaics, Empa–Swiss Federal Laboratories for Materials Science and Technology, CH-8600 Duebendorf, Switzerland. [10]Molecular Chirality Research Center, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan.
***Corresponding author. Email: feng.wang@liu.se (F.W.); feng.gao@liu.se (F.G.)**
†These authors contributed equally to this work.

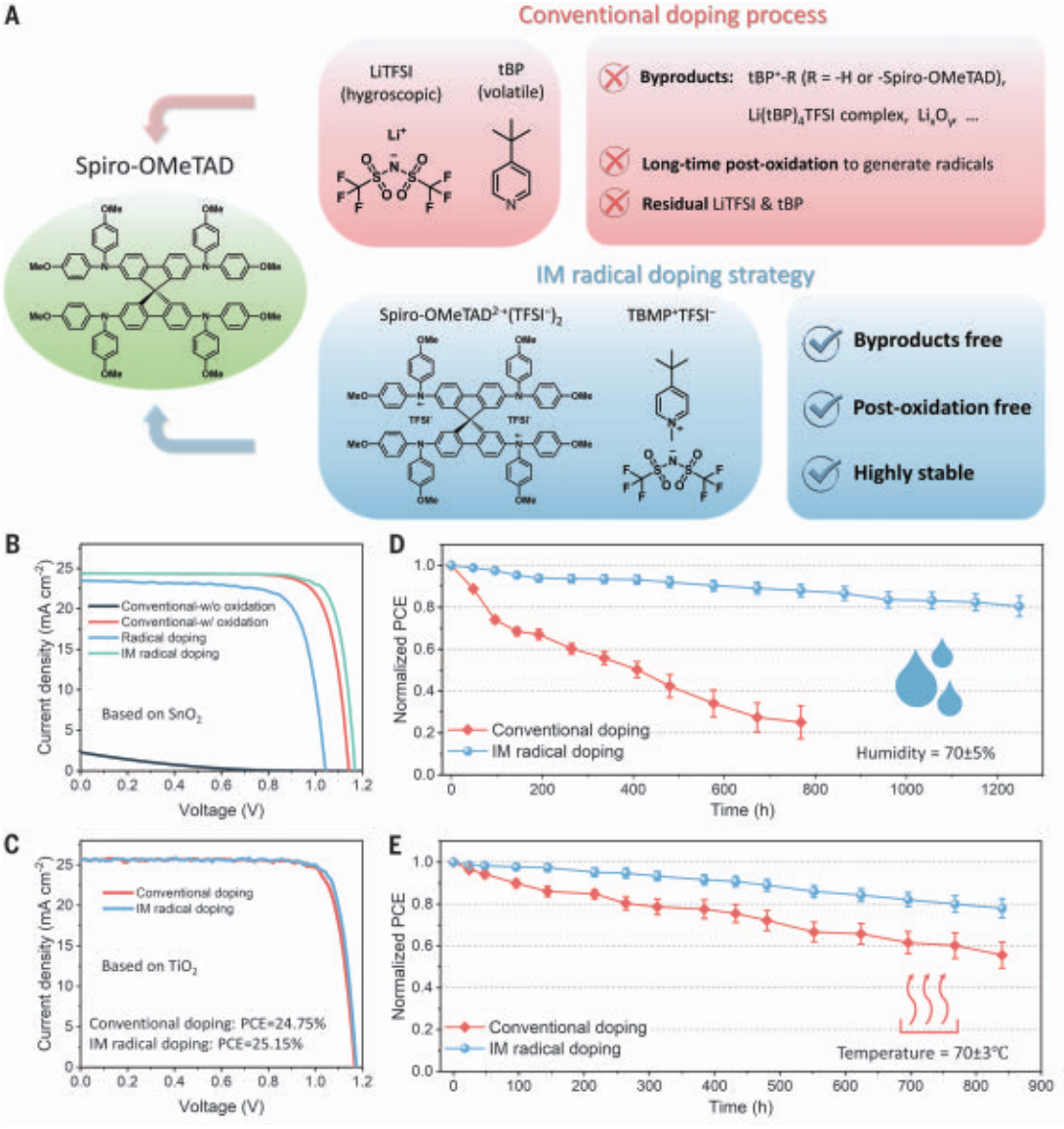


Fig. 1. Comparison of perovskite solar cells (PSCs) based on the conventional and ion-modulated (IM) radical doping strategies. (**A**) Illustration of the complex in situ reaction processes in the conventional doping process (top) and the clean, instant IM radical doping strategy (bottom) of spiro-OMeTAD. The radical and ionic salt were dissolved in 1,1,2,2-tetrachloroethane. (**B**) Current density–voltage (*J-V*) curves of PSCs (SnO_2 electron transport layer) based on conventional doping, radical doping, and IM radical doping of spiro-OMeTAD. (**C**) *J-V* curves of PSCs (mesoporous TiO_2 electron transport layer) based on conventional and IM radical doping of spiro-OMeTAD. (**D** and **E**) PCE tracking of unencapsulated PSCs based on conventional and IM radical doping of spiro-OMeTAD under 70 ± 5% humidity (D) and 70 ± 3°C thermal aging (E). Error bars denote SD.

acts as the main dopant to improve conductivity and WF. The second is the ionic salt 4-*tert*-butyl-1-methylpyridinium bis(trifluoromethylsulfonyl)imide ($TBMP^+TFSI^-$) to further modulate the WF. In this doping strategy, the localized ionic environment created by the ionic salt was used to manipulate the energetics of the hole polarons and decouple the conductivity and WF tunability. The doped spiro-OMeTAD based on our IM radical doping strategy delivered PSCs simultaneously with high efficiency (PCE > 25%) and high stability (T_{80} for ~1200 hours under 70 ± 5% relative humidity (RH) and T_{80} for ~800 hours under 70° ± 3°C without encapsulation), minimizing the trade-off between efficiency and stability of PSCs. In addition, the IM radical doping strategy provides a facile yet effective approach to separately optimize the conductivity and WF of organic semiconductors for a variety of optoelectronic applications.

PSC performance based on IM radical–doped spiro-OMeTAD

We first compared the distinctive effects of conventional and IM radical doping of spiro-OMeTAD using $(FAPbI_3)_{0.99}(MAPbBr_3)_{0.01}$ (FA, formamidinium; MA, methylammonium) PSCs with an SnO_2 electron transport layer (ETL) (fig. S3) (*12*). The conventional spiro-OMeTAD doping process requires an oxidation time of ~24 hours to reach the optimized PCE of 22.2% (Fig. 1B, fig. S4A, and table S1). The gradual improvement in device performance with increasing oxidation time is consistent with the conductivity increase (fig. S5A) that resulted from the continuous generation of spiro-$OMeTAD^{\bullet+}TFSI^-$ radicals, as evidenced by the ultraviolet–visible–near-infrared (UV-Vis-NIR) absorption spectra (fig. S5B) (*29*, *30*).

In our IM spiro-OMeTAD doping strategy, the first additive consists of presynthesized spiro-$OMeTAD^{2\bullet+}(TFSI^-)_2$ (*31*, *32*), which can be immediately converted into spiro-$OMeTAD^{\bullet+}TFSI^-$ monoradicals through comproportionation with the neutral spiro-OMeTAD (*25*):

$$\text{spiro-OMeTAD} + \text{spiro-OMeTAD}^{2\bullet+}(TFSI^-)_2 \rightarrow 2\ \text{spiro-OMeTAD}^{\bullet+}TFSI^- \quad (1)$$

We note that more radical cations (and hence a higher conductivity) are obtained at lower dopant levels using the current approach than using LiTFSI, as indicated by absorption spectra and conductivity measurements (fig. S5). Upon the incorporation of radicals, the device performance improved relative to undoped spiro-OMeTAD (fig. S4B), and an optimal performance was reached with the doping radical ratio of ~14 mol % (calculated to the ratio of monoradicals). However, compared to a high

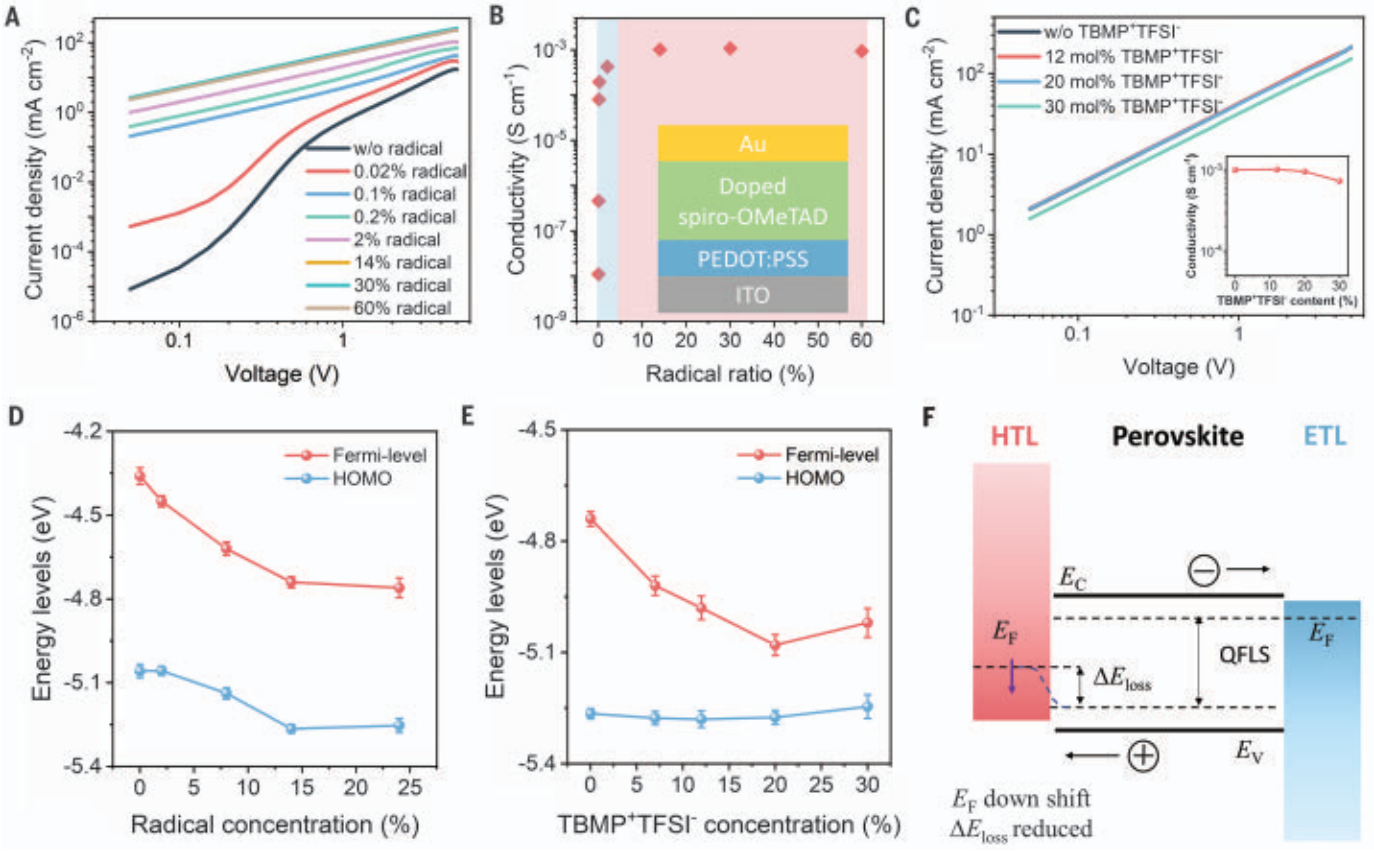


Fig. 2. Effects of radicals and ionic salts on conductivity and energetics. (**A** and **B**) *J-V* curves of the hole-only devices (A) and conductivity of the spiro-OMeTAD films doped with different radical amounts (B). Inset of (B) shows the structure of the hole-only devices. (**C**) *J-V* curves of the hole-only devices with 14 mol % radicals and different amounts of TBMP$^+$TFSI$^-$. The inset shows conductivity. (**D** and **E**) Fermi-level and HOMO onsets of the spiro-OMeTAD films doped with different radical amounts (D) and different TBMP$^+$TFSI$^-$ amounts (14 mol % radicals) (E). Error bars denote SD. (**F**) Illustration of the band alignment between the perovskite layer and the HTL with different WF values.

open-circuit voltage (V_{oc}) of 1.15 V achieved with the conventional doping recipe, the optimized V_{oc} of spiro-OMeTAD doped with radicals was 1.04 V (Fig. 1B).

Upon addition of the second additive (ionic salt, TBMP$^+$TFSI$^-$), the V_{oc} of the solar cells increased and a V_{oc} of 1.17 V was achieved with 20 mol % TBMP$^+$TFSI$^-$. The resulting PCE of 23.4 % was even higher than that based on the conventional doping (Fig. 1B). Further increasing the TBMP$^+$TFSI$^-$ content led to a decrease in the fill factor (FF) (fig. S4C and table S2) possibly because of ion aggregation (fig. S6).

The high V_{oc} of the optimized devices was also consistent with the values of electroluminescence external quantum efficiency (EL_{EQE}), which increased from 0.05% to 5.4% upon the addition of 20 mol % TBMP$^+$TFSI$^-$ into the HTL (fig. S7), indicating reduced nonradiative recombination. Remarkably, based on the meso-TiO_2 scaffold with $FAPbI_3$ as the active layer, the IM radical doping of spiro-OMeTAD delivered a PCE of 25.1% (table S3, certified 25.0% in fig. S8), which was again higher than that of the control device with conventionally doped spiro-OMeTAD (Fig. 1C). The PCE statistics demonstrated that the IM radical doping strategy, which did not involve a post-oxidation process, was more controllable than the conventional doping process (fig. S9).

Because the IM radical doping strategy completely removed the need to use hygroscopic LiTFSI and volatile tBP, an improvement in device stability can be expected. We systematically compared the device stability with IM radical doping versus conventional doping as a function of RH, heat, and illumination. Relative to the conventional doping of spiro-OMeTAD, the IM radical doping strategy extended the T_{80} of PSCs (unencapsulated) from ~96 to ~1240 hours under high RH of ~70 ± 5% (Fig. 1D and fig. S10), and T_{80} from ~264 to ~796 hours under ~70° ± 3°C (Fig. 1E and fig. S11). The improved device stability with the IM radical–doped HTL was ascribed to not only the improved stability of HTL but also the prevention of phase and thermal degradation of the perovskite active layer under high RH and under heat stress, as evidenced from optical images (fig. S12) and x-ray diffraction results (figs. S13 and S14) of the degraded devices.

Scanning electron microscopy (SEM) images provided more details of the protection effect from the spiro-OMeTAD HTL based on IM radical doping. A large number of pinholes were observed in degraded devices based on the conventionally doped HTLs, implying that LiTFSI residuals absorbed moisture and that tBP volatilized gradually. Both effects accelerated the degradation of the perovskite active layer (figs. S15 and S16). In contrast, the devices based on the IM radical doping strategy maintained a compact and uniform morphology after long-term exposure to high RH and thermal stress (figs. S15 and S17), protecting the perovskite layer underneath. In addition, under maximum power point (MPP) tracking, the PCEs of PSCs based on conventional doping decayed to ~85% after 500 hours of continuous illumination. Devices based on the IM radical doping strategy maintained ~95% of the initial PCE after the same operational time (fig. S18), similar to that of devices using inorganic HTLs (*33*, *34*) or undoped polymer HTLs (*35*).

The roles of radicals and ionic salts on spiro-OMeTAD doping

Our IM radical doping strategy provided an instantly effective and stable doping of spiro-OMeTAD. In addition, this simple process

allowed us to decouple and investigate the fundamental functions of additives (radicals and ionic salts). We focused on two physical parameters that are key to the effectiveness of the HTL: conductivity and energy levels [including both WF and ionization energy (IE)].

To investigate the HTL conductivity, we used hole-only devices with the structure ITO/PEDOT:PSS/doped spiro-OMeTAD/Au [ITO, indium tin oxide; PEDOT, poly(3,4-ethylenedioxythiophene); PSS, polystyrene sulfonate]. The effects of organic radicals on the film conductivity (Fig. 2, A and B) showed two distinct stages as a function of an increasing amount of organic radicals. There was initially an extremely high doping efficiency with a small doping ratio of radicals (<2 mol %; Fig. 2B, light blue area), which suggested efficient hole polaron generation. In particular, 0.1 mol % of radicals could increase the conductivity of the film by almost four orders of magnitude, from 1.10×10^{-8} S cm^{-1} to 8.03×10^{-5} S cm^{-1}. This increase could be attributed to the filling of deep trap states by dopant-generated charge carriers, which would bring increased carrier concentration as well as higher carrier mobility (*36*).

For higher doping concentrations, the hole-only current density exhibited a linear dependence over the entire voltage range from 0 V to 5 V, indicative of a high hole concentration in the system. In the second stage (doping ratio from 2 mol % to 60 mol %, marked by the light red area), the doping efficiency nearly saturated and the conductivity peaked at ~1.00×10^{-3} S cm^{-1} at ~14 mol % radicals (Fig. 2B). This doping saturation phenomenon was consistent with previous reports in other organic semiconductor systems and is ascribed to the increase in energetic disorder induced by long-range Coulomb interactions when increasing the number of dopants (*37*). The addition of the $TBMP^+TFSI^-$ ionic salt had negligible effects on the film conductivity over a wide range of concentrations (up to 20 mol %) (Fig. 2C), beyond which a slight decrease in film conductivity was observed, possibly because of the aforementioned phase aggregation and increased disorder in the film (see below).

In addition to conductivity, the energy level alignment at the HTL interface is another parameter that is critically important to device performance. An optimal energy level of the HTL helps to facilitate hole extraction and eliminate interfacial nonradiative recombination, contributing to high device performance (especially high V_{oc}). We investigated the impact of the radicals on the relevant energy levels by ultraviolet photoelectron spectroscopy (UPS). For the pristine spiro-OMeTAD, the ionization energy (IE) was 5.08 eV, which according to Koopmans' theorem (*38*) corresponds to the energy of the highest occupied molecular orbital (HOMO), and the WF was 4.36 eV (fig. S19A). The density functional theory (DFT) results showed that the IE of an isolated spiro-OMeTAD molecule was ~5.46 eV, and this value decreased to 5.03 eV when the calculations were performed by taking into account an effective medium with a dielectric constant of 2.4 (fig. S20).

With increasing amounts of radicals, both IE and WF increased (Fig. 2D). A more detailed comparison revealed that the difference between WF and IE decreases along with increasing dopant concentration (Fig. 2D), emphasizing the extent to which the Fermi level moves toward the HOMO level, which is a key signature of p-doping. Combining the conductivity results in Fig. 2B, we noticed that a radical concentration of >2 mol % could affect the energy levels of the film despite almost no improvement in conductivity. The shift of energetic levels saturated at ~14 mol % radical doping, which is due to the limit in charge transfer and hole generation (*37*).

Based on the 14 mol % radical–doped system, we further assessed the effect of the $TBMP^+TFSI^-$ ionic salt on the film energy levels. In contrast to the simultaneous change of the HOMO level and WF upon the addition of radicals, the ionic salt increased the WF to 5.08 eV with 20 mol % $TBMP^+TFSI^-$ but had a negligible effect on the HOMO levels (Fig. 2E and fig. S19B). The slightly decreased WF with 30% $TBMP^+TFSI^-$ was ascribed to the effects of ion aggregation (fig. S6). As such, the addition of $TBMP^+TFSI^-$ breaks the doping limit of radical doping, resulting in a Fermi level that is very close to the HOMO level (the difference is as small as 0.12 eV, a value much smaller than those obtained in traditional organic semiconductor doping). The promoted doping efficiency after $TBMP^+TFSI^-$ incorporation was also confirmed by the increased carrier concentration, as indicated from a Mott-Schottky analysis (fig. S21).

In PSCs, as long as the HOMO energy of the HTL remains above the valence band of perovskites, it has little impact on the device V_{oc} value. For example, different HTLs with HOMO energies between −5.2 and −5.4 eV have almost the same *J-V* curves (*39*). Within a suitable HOMO energy range, an optimal WF of the HTL aligned with the quasi–Fermi level (QFL) of the perovskite active layer is critical to minimizing the voltage loss (*40*, *41*). As illustrated in Fig. 2F, the increase of the WF by the $TBMP^+TFSI^-$ reduced the hole extraction barrier between the perovskite and the IM radical–doped HTL film. This change enables a high QFL splitting in the perovskites under illumination by eliminating the interfacial nonradiative recombination loss. Thus, this WF modulation by the ionic salt is important, as this V_{oc} improvement made it possible for the device based on our IM radical doping strategy to reach the performance comparable to that based on conventional doping (refer to Fig. 1, B and C).

From these conductivity and energy level investigations, the different functions assumed by the organic radicals and ionic salts become clear. The radicals enhance the conductivity and affect the energy levels (both HOMO and WF); the ionic salts further optimize the WF and contribute to optimal V_{oc} values in the devices.

The effectiveness of the radicals as a means of tuning the conductivity and energy levels can be understood with a charge transfer model between neutral spiro-OMeTAD and the spiro-$OMeTAD^{\bullet+}TFSI^-$ radical. For neutral spiro-OMeTAD, the Fermi level (E_F) corresponds to an energy about midway between the lowest unoccupied molecular orbital (LUMO) and HOMO levels (fig. S20). The low film conductivity in the absence of dopant can be understood by poor film crystallinity and low charge carrier concentrations, which are determined by thermal excitation of charge carriers from HOMO to LUMO levels.

Upon radical doping, spiro-$OMeTAD^{\bullet+}TFSI^-$ provides additional electronic states (*42*) (figs. S22 to S24)—specifically, bound hole states corresponding to singly unoccupied molecular orbital (SUMO) levels stabilized by Coulombic attraction from the $TFSI^-$ counterion. Electron transfer from neutral spiro-OMeTAD (host) to spiro-$OMeTAD^{\bullet+}TFSI^-$ radical complex (dopant) leads to the generation of free hole polarons and increases the conductivity (Fig. 3A). Concurrently, the hole filling of the transport states downshifts the Fermi level and increases the WF (Fig. 2D) (*43*, *44*).

The mechanism of ionic salts modulating spiro-OMeTAD doping

We now turn to a discussion of how ionic salts can further tune the WF. We did not observe oxidation or reduction between ionic salts and neutral spiro-OMeTAD/spiro-$OMeTAD^{\bullet+}TFSI^-$ (figs. S25 and S26). A straightforward possibility for inducing the WF change is the introduction of a macroscopic dipole, which could be induced either by a particular spiro-OMeTAD orientation (*45*) or by an interfacial $TBMP^+TFSI^-$ accumulation (*46*). We excluded the former on the basis of the grazing-incidence wide-angle x-ray scattering (GIWAXS) results (fig. S27), which indicated neither stacking mode change nor crystallization amelioration upon addition of $TBMP^+TFSI^-$. The latter could be ruled out because time-of-flight secondary ion mass spectroscopy (TOF-SIMS) measurements showed a uniform distribution of S^-, $C_4H_9^-$, $C_5H_4N^-$, and CF_3^- across the entire HTL (fig. S28), indicating the absence of surface dipoles formed by $TBMP^+TFSI^-$. In addition, the incorporation of ionic salts increased the WF more than its effects on the IE (Fig. 2D), further ruling out the possibility of the surface-dipole effect.

The exclusion of macroscopic dipole formation motivated us to understand the WF tuning

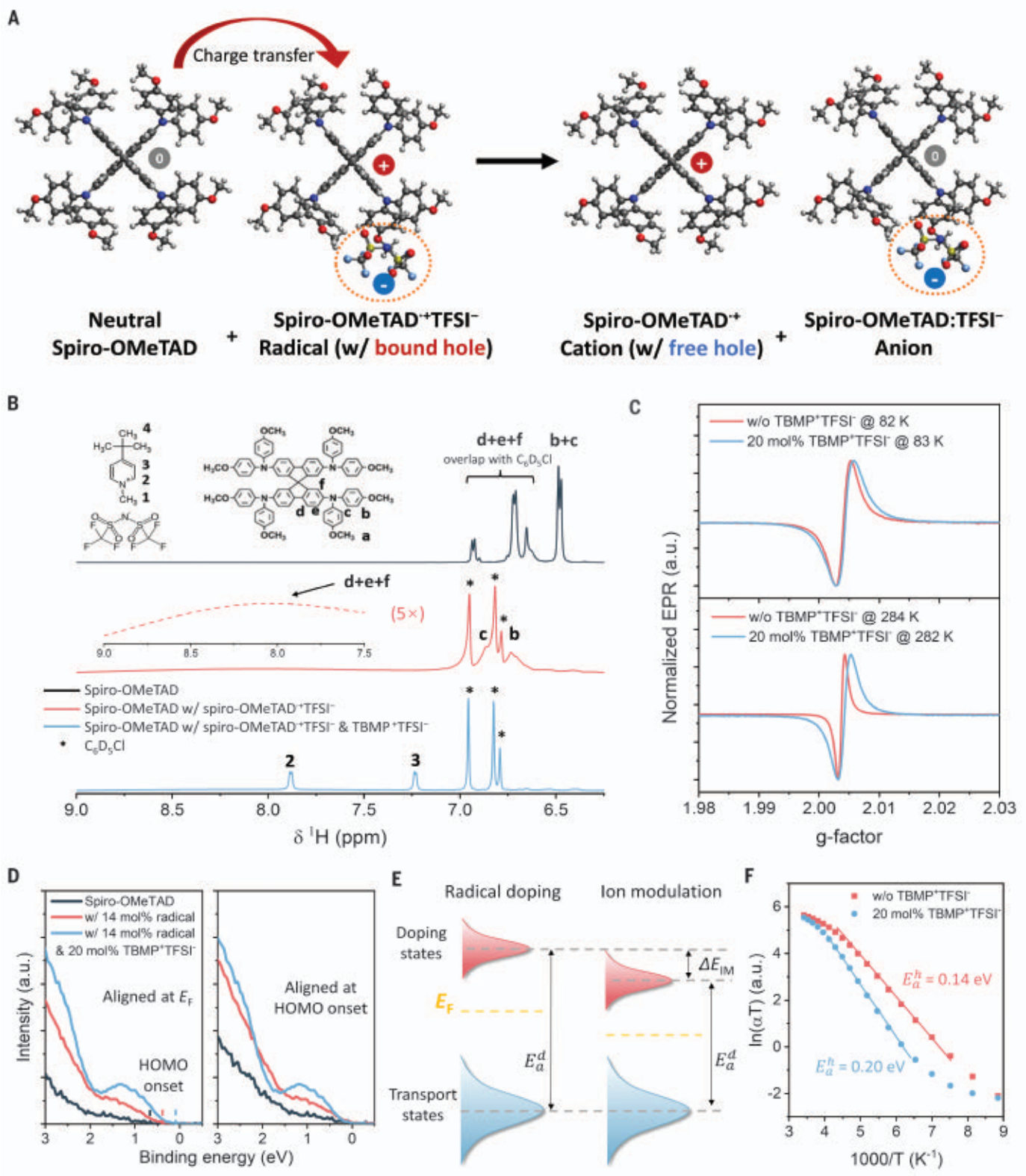


Fig. 3. Molecular-level doping mechanisms of the IM radical doping. (**A**) Illustration of the charge transfer and doping mechanisms with spiro-OMeTAD$^{\cdot+}$TFSI$^-$ radicals. (**B**) ^{1}H hr-NMR spectra of spiro-OMeTAD (black line), spiro-OMeTAD/spiro-OMeTAD$^{\cdot+}$TFSI$^-$ mixture without TBMP$^+$TFSI$^-$ (red line; inset shows magnified view of d+e+f peak in the range of 9 to ~7.5 ppm), and spiro-OMeTAD/spiro-OMeTAD$^{\cdot+}$TFSI$^-$ mixture with TBMP$^+$TFSI$^-$ (blue line) in the range 9 to 6.25 ppm. (Peaks b, c, and d+e+f refer to the aromatic protons signal.) (**C**) EPR signals of doped spiro-OMeTAD (14 mol % radicals) with and without TBMP$^+$TFSI$^-$ at low temperature and room temperature. (**D**) Zoom-in spectra near the valence band of neutral spiro-OMeTAD, 14 mol % radical–doped spiro-OMeTAD film, and 14 mol % radical and 20 mol % TBMP$^+$TFSI$^-$–doped spiro-OMeTAD films. (**E**) Illustration of ionic salts' effect on WF modulation in the framework of the IM radical doping strategy. (**F**) Temperature-dependent conductivity evolution of doped spiro-OMeTAD (14 mol % radicals) with and without TBMP$^+$TFSI$^-$.

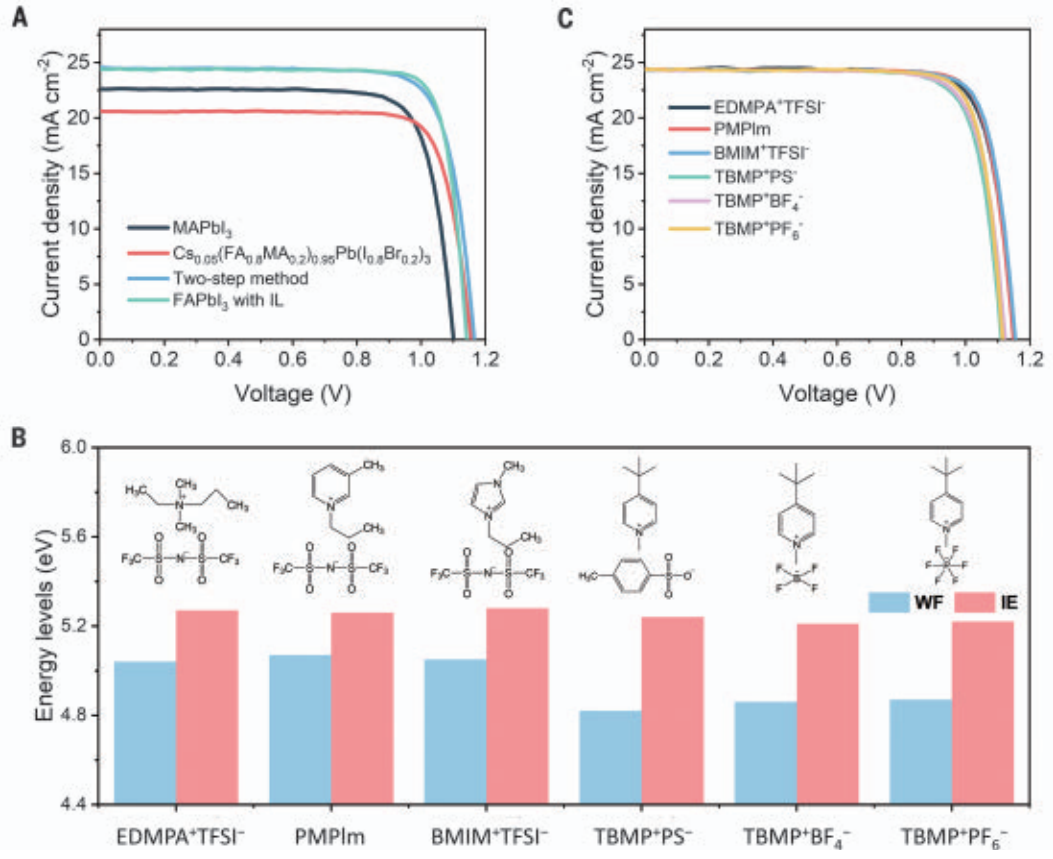


Fig. 4. Generality of the IM radical doping strategy. (**A**) *J-V* curves of different PSCs based on spiro-OMeTAD HTL following the IM radical doping strategy. (**B**) Molecular structures of six additional ionic salts used in this work and their effects on the energetic levels of the doped spiro-OMeTAD films. EDMPA, ethyldimethylpropylammonium; PMPIm, 3-methyl-1-propylpyridinium; BMIM, 1-propyl-3-methyl-imidazolium. (**C**) *J-V* curves of PSCs based on the spiro-OMeTAD HTL doped with six additional ionic salts (14 mol % radicals).

mechanism at the molecular level. Both radicals and salts are ionic, so Coulomb interactions among them are to be expected and would affect their frontier orbital energetics. Details on the Coulomb interactions between the ionic salts and radicals were obtained from both high-resolution liquid and solid-state nuclear magnetic resonance (hr- and ss-NMR, respectively) spectroscopies (Fig. 3B). Although NMR spectroscopy cannot provide direct information on radicals given the strong interactions of unpaired electrons with NMR active nuclei, comparison between different samples provided useful information on the interactions in our mixture. We observed electron transfer between spiro-OMeTAD$^{\bullet+}$TFSI$^-$ and spiro-OMeTAD, in that all ^{1}H signals (labeled a to f) of spiro-OMeTAD hr-NMR spectra were broadened after the addition of spiro-OMeTAD$^{\bullet+}$TFSI$^-$ (Fig. 3B and fig. S29). The broad signal half-widths suggested that spiro-OMeTAD$^{\bullet+}$TFSI$^-$ and neutral spiro-OMeTAD were near each other and mutually exchanging electrons, which could be attributed to π-π stacking between fluorene moieties of spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ (*47*, *48*). However, the ^{1}H NMR signals of spiro-OMeTAD d+e+f were more broadened while the b and c peaks were only slightly affected, indicating that the electron exchange between spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ influenced the molecules only partially on the fluorene part (*49*).

Upon addition of TBMP$^+$TFSI$^-$ into the spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ mixture, the ^{1}H NMR signals from all aromatic protons (b, c, and d+e+f) of spiro-OMeTAD nearly disappeared (Fig. 3B). This indicated more effective electron exchange between spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ in the presence of the TBMP$^+$TFSI$^-$ salt. Because direct interactions between TBMP$^+$TFSI$^-$ and spiro-OMeTAD were negligible (fig. S29A), we attributed the effect of TBMP$^+$TFSI$^-$ on the ^{1}H hr-NMR signals (Fig. 3B) to Coulomb interactions between spiro-OMeTAD$^{\bullet+}$TFSI$^-$ and TBMP$^+$TFSI$^-$, which enhanced charge transfer between spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ in the presence of the TBMP$^+$TFSI$^-$ salt. The emergence of these interactions is consistent with the shift and narrowing of the ^{19}F ss-NMR peak of TFSI$^-$ (fig. S30).

A complementary electron paramagnetic resonance (EPR) study from 77 K to room temperature demonstrated a Lorentzian broadening of EPR spectra in our radical-doped spiro-OMeTAD (for samples both with and without TBMP$^+$TFSI$^-$) (Fig. 3C and figs. S31 and S32). We attributed this effect to the dynamic broadening due to the electron self-exchange between the neutral spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ radical. Within the intermediate exchange rate limit, the EPR broadening upon TBMP$^+$TFSI$^-$ incorporation was attributed to the increased electron transfer rate between the neutral spiro-OMeTAD and spiro-OMeTAD$^{\bullet+}$TFSI$^-$ radical that decreased the spin lifetime (*50*).

We further investigated the impact of these Coulomb interactions on the frontier orbital energies by combining the photoelectron spectroscopy data and DFT results. The incorporation of TBMP$^+$TFSI$^-$ not only decreased the energy offset between the HOMO level and the Fermi level but also changed the density of states (DOS) of frontier orbitals (Fig. 3D). In addition, low-energy inverse photoelectron spectra (LEIPS) indicated that the SUMO level of spiro-OMeTAD$^{\bullet+}$ downshifted from −4.06 to −4.30 eV upon addition of TBMP$^+$TFSI$^-$ (fig. S33). The DFT calculations highlighted the effect of the ionic salt on the molecular orbital levels (figs. S34 and S35), which we rationalized by considering that the TBMP$^+$TFSI$^-$ ionic salt screened the Coulomb attraction between the counterion and hole polaron in the spiro-OMeTAD$^{\bullet+}$TFSI$^-$ complex (*51*). In this context, the impact of the TBMP$^+$TFSI$^-$ can be viewed as a modulation of the activation energy of doping (E_a^d) related to the energy difference between the doping states and transport states (Fig. 3E), because neutral spiro-OMeTAD molecules close to TFSI$^-$ anions from the ionic salts can homogenize the electrostatic potential of the transport states, resulting in decreased E_a^d and increased WF.

Given this understanding of the impact of ionic salts on the energetic levels, we investigated

why there was no conductivity enhancement with increasing WF, a feature that makes it decouple conductivity and WF tuning. Decreasing the E_a^d of dopants upon ionic salt addition (Fig. 3E) enabled an increasing fraction of hole polarons to be thermally populated. This effect, taken alone, would lead to a conductivity increase. However, the temperature-dependent conductivity indicated that the ionic salts also increase disorder. The temperature-dependent conductivity measurements revealed a faster-decreasing rate of conductivity upon addition of $TBMP^+TFSI^-$ and suggested an increased long-distance intermolecular charge hopping potential barrier (E_a^h) from 0.14 eV to 0.20 eV (*52*, *53*) (Fig. 3F and fig. S36). This more difficult long-distance intermolecular charge transfer implied a decreased mobility that might be induced by increased disorder or more scattering from ionic clusters (*54*, *55*). Overall, the ionic salts played a dual role: They increased the free hole population but also increased disorder. These two effects largely compensated each other and led to negligible conductivity changes over a broad range of ionic salt concentrations.

The generality of the IM radical doping strategy

Stimulated by the possibilities of precise energetic manipulation and facile conductivity tunability, we applied the IM radical doping approach to perovskites with different compositions and devices with different architectures (planar or mesoporous structures). In addition to the PSCs based on SnO_2/$(FAPbI_3)_{0.99}(MAPbBr_3)_{0.01}$ and TiO_2/$FAPbI_3$ shown in Fig. 1, B and C, IM radical–doped spiro-OMeTAD HTLs also resulted in high PCEs in perovskites with different band gaps, different fabrication methods (one-step or two-step), and with or without ionic liquid additives. All PCEs were similar to (or slightly higher than) those based on conventionally doped spiro-OMeTAD HTL (Fig. 4A, fig. S37, and tables S4 and S5), indicating that our IM radical doping strategy is applicable to different PSCs. In addition, we observed improved stability (both under high RH and high temperature) for all the PSCs using IM radical–doped spiro-OMeTAD compared with those using conventionally doped spiro-OMeTAD (fig. S38).

The generality of the IM radical doping strategy was also demonstrated in terms of the ionic salts. We investigated six additional ionic salts, based on typical anions [bis(trifluoromethylsulfonyl)imide, hexafluorophosphate, tetrafluoroborate, p-toluenesulfonate] and cations (ethyldimethylpropylammonium, propylpyridinium, imidazolium, methylpyridinium) (Fig. 4B). All of the ionic salts could effectively tune the WF (Fig. 4B and fig. S39) without negatively affecting the conductivity (fig. S40). Their incorporation resulted in high V_{oc} values (~1.12 to 1.16 V) and high PCEs (~22% to 23%) in SnO_2/$(FAPbI_3)_{0.99}(MAPbBr_3)_{0.01}$-based PSCs (Fig. 4C and table S6). Although the use of different cations resulted in negligible WF modifications, we noted that switching from $TFSI^-$ to other anions could slightly decrease the WF, which pointed to the anions playing a more prominent role in the energetic modulation in the framework of the IM radical doping strategy. A possible reason is that the anions coming from the ionic salts can exchange with $TFSI^-$ in the radical complex, resulting in new radical complex species with different energetics.

Discussion

Because the energetics of the hole polarons are sensitive to their local ionic environment, the IM radical doping strategy provides room to manipulate the WF as a function of the nature of the ionic salts. WF tuning could be achieved over a large range by controlling the interaction strength between the ionic salts and the radical complex and is more easily adjustable than WF modulation through interfacial molecular dipoles (*46*). As such, the IM radical doping strategy fills the gap between several previous strategies to enhance WF and conductivity and greatly expands the toolbox for organic semiconductor doping (*56*, *57*).

REFERENCES AND NOTES

1. H. Min *et al*., *Nature* **598**, 444–450 (2021).
2. T. Zhang *et al*., *Joule* **2**, 2706–2721 (2018).
3. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
4. S. Chen *et al*., *Nat. Sustain.* **4**, 636–643 (2021).
5. J. Xue *et al*., *Science* **371**, 636–640 (2021).
6. Y. H. Lin *et al*., *Science* **369**, 96–102 (2020).
7. A. Al-Ashouri *et al*., *Science* **370**, 1300–1309 (2020).
8. K. Xiao *et al*., *Nat. Energy* **5**, 870–880 (2020).
9. J. Jeong *et al*., *Nature* **592**, 381–385 (2021).
10. M. Jeong *et al*., *Science* **369**, 1615–1620 (2020).
11. J. J. Yoo *et al*., *Nature* **590**, 587–593 (2021).
12. M. Kim *et al*., *Joule* **3**, 2179–2192 (2019).
13. G. Kim *et al*., *Science* **370**, 108–112 (2020).
14. S. Wang *et al*., *Nano Lett.* **16**, 5594–5600 (2016).
15. U. B. Cappel, T. Daeneke, U. Bach, *Nano Lett.* **12**, 4925–4931 (2012).
16. R. L. Forward *et al*., *ACS Energy Lett.* **4**, 2547–2551 (2019).
17. S. G. Kim *et al*., *Adv. Mater.* **33**, e2007431 (2021).
18. A. K. Jena, M. Ikegami, T. Miyasaka, *ACS Energy Lett.* **2**, 1760–1761 (2017).
19. A. K. Jena, Y. Numata, M. Ikegami, T. Miyasaka, *J. Mater. Chem. A* **6**, 2219–2230 (2018).
20. J. Kong *et al*., *Nature* **594**, 51–56 (2021).
21. B. Xu *et al*., *ChemSusChem* **7**, 3252–3256 (2014).
22. J.-Y. Seo *et al*., *Energy Environ. Sci.* **11**, 2985–2992 (2018).
23. J.-Y. Seo *et al*., *Adv. Func. Mater.* **31**, 2102124 (2021).
24. A. Abate *et al*., *J. Am. Chem. Soc.* **135**, 13538–13548 (2013).
25. W. H. Nguyen, C. D. Bailie, E. L. Unger, M. D. McGehee, *J. Am. Chem. Soc.* **136**, 10996–11001 (2014).
26. B. Tan *et al*., *Adv. Energy Mater.* **9**, 1901519 (2019).
27. S. Wang *et al*., *J. Am. Chem. Soc.* **140**, 16720–16730 (2018).
28. F. Lamberti *et al*., *Chem* **5**, 1806–1817 (2019).
29. S. Fantacci, F. De Angelis, M. K. Nazeeruddin, M. Grätzel, *J. Phys. Chem. C* **115**, 23126–23133 (2011).
30. V. Coropceanu, M. Malagoli, J. M. André, J. L. Brédas, *J. Am. Chem. Soc.* **124**, 10519–10530 (2002).
31. W. Zhang *et al*., *Chem. Commun.* **56**, 1589–1592 (2020).
32. W. Zhang *et al*., *ACS Appl. Mater. Interfaces* **12**, 33751–33758 (2020).
33. J. You *et al*., *Nat. Nanotechnol.* **11**, 75–81 (2016).
34. W. Chen *et al*., *Science* **350**, 944–948 (2015).
35. Y. Hou *et al*., *Science* **358**, 1192–1197 (2017).
36. S. Olthof *et al*., *Phys. Rev. Lett.* **109**, 176601 (2012).
37. M. L. Tietze, P. Pahner, K. Schmidt, K. Leo, B. Lüssem, *Adv. Funct. Mater.* **25**, 2701–2707 (2015).
38. T. Koopmans, *Physica* **1**, 104–113 (1934).
39. R. A. Belisle, P. Jain, R. Prasanna, T. Leijtens, M. D. McGehee, *ACS Energy Lett.* **1**, 556–560 (2016).
40. P. Caprioglio *et al*., *Adv. Energy Mater.* **9**, 1901631 (2019).
41. M. Stolterfoht *et al*., *Energy Environ. Sci.* **12**, 2778–2788 (2019).
42. A. Fediai, F. Symalla, P. Friederich, W. Wenzel, *Nat. Commun.* **10**, 4547 (2019).
43. C. G. Tang *et al*., *Nature* **539**, 536–540 (2016).
44. Y. Yamashita *et al*., *Nature* **572**, 634–638 (2019).
45. N. Shibayama *et al*., *ACS Appl. Mater. Interfaces* **12**, 50187–50191 (2020).
46. Y. Zhou *et al*., *Science* **336**, 327–332 (2012).
47. Y. Li, H. Li, C. Zhong, G. Sini, J.-L. Brédas, *Npj Flex. Electron.* **1**, 1–8 (2017).
48. D. Shi *et al*., *Sci. Adv.* **2**, e1501491 (2016).
49. E. de Boer, C. MacLean, *J. Chem. Phys.* **44**, 1334–1342 (1966).
50. R. L. Ward, S. I. Weissman, *J. Am. Chem. Soc.* **76**, 3612 (1954).
51. R.-Q. Png *et al*., *Nat. Commun.* **7**, 11948 (2016).
52. N. Karl, *Synth. Met.* **133**, 649–657 (2003).
53. H. Abdalla, G. Zuo, M. Kemerink, *Phys. Rev. B* **96**, 241202 (2017).
54. P.-J. Chia *et al*., *Phys. Rev. Lett.* **102**, 096602 (2009).
55. K. Harada, F. Li, B. Maennig, M. Pfeiffer, K. Leo, *Appl. Phys. Lett.* **91**, 092118 (2007).
56. I. E. Jacobs, A. J. Moulé, *Adv. Mater.* **29**, 1703063 (2017).
57. C. Gaul *et al*., *Nat. Mater.* **17**, 439–444 (2018).

ACKNOWLEDGMENTS

We thank C. Bao for the help with temperature-dependent conductivity measurements; G. Zheng for help with the GIWAXS analysis; Y. Wang for help with the stability test; J. C. M. Yu for help with single-crystal characterization; and O. Inganäs, W. Xu, X. Liu, and X. Cai for valuable discussions. **Funding:** Supported by Swedish Research Council Vetenskapsrådet (grant 2018-04809), an ERC Starting Grant (717026), the Knut and Alice Wallenberg Foundation (Dnr KAW 2019.0082), and the Swedish Government Strategic Research Area in Materials Science on Functional Materials at Linköping University (Faculty Grant SFO-Mat-LiU 2009-00971). F.G. is a Wallenberg Academy Fellow. Work at the University of Arizona was funded by the UA College of Science and the Office of Naval Research, awards N00014-20-1-2110 and N00014-22-1-2379. Work at the École Polytechnique Fédérale de Lausanne was funded by Swiss National Science Foundation for financial support with project 200020_185041 and from the European Union's Horizon 2020 research and innovation program under grant agreement 764047. Work at Chiba University was funded by Advanced Low Carbon Technology Research and Development Program (ALCA) of Japan Science and Technology Agency (grant JPMJAL 1404). **Author contributions:** F.W. and F.G. conceived and supervised the project. T.Z., Z.H., F.W., and F.G. designed this project. Z.H. contributed to advising and providing chemistry solutions. T.Z., Z.H., and F.W. performed the investigations and analyzed the data. H.-B.K., I.W.C., and D.S.K. made the PSCs based on the meso-TiO_2. C.W., X.L., and M.F. conducted XPS and UPS measurements and analysis. E.C., V.C., and J.-L.B. conducted the DFT calculations and helped to discuss the doping mechanism. R.K., L.K., and J.B. performed NMR measurements. M.C. performed SEM measurements and helped with NMR interpretations. M.Y. and S.B. contributed to the discussions. Y.P. and W.M.C. performed EPR analysis. B.Y., J.S., and A.H. performed the MPP measurements. K.T. and H.Y. conducted the LEIPS measurements. S.-C.Y. and F.F. contributed to TOF-SIMS measurements. T.Z., F.W., and F.G. wrote the manuscript. All authors discussed the results and reviewed the manuscript. **Competing interests:** F.G., F.W., T.Z., and Z.H. have filed a patent based on the concept developed in this manuscript. They are also co-founders of the company LinXole AB, which commercializes the materials developed in this manuscript. The other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo2757
Materials and Methods
Supplementary Text
Figs. S1 to S40
Tables S1 to S6
References (*58*–*66*)

Submitted 24 January 2022; accepted 22 June 2022
10.1126/science.abo2757

CANCER

Mechanism-based design of agents that selectively target drug-resistant glioma

Kingson Lin[1,2,3]†, Susan E. Gueble[2]†, Ranjini K. Sundaram[2], Eric D. Huseman[1], Ranjit S. Bindra[2,3]*, Seth B. Herzon[1,4]*

Approximately half of glioblastoma and more than two-thirds of grade II and III glioma tumors lack the DNA repair protein O^6-methylguanine methyl transferase (MGMT). MGMT-deficient tumors respond initially to the DNA methylation agent temozolomide (TMZ) but frequently acquire resistance through loss of the mismatch repair (MMR) pathway. We report the development of agents that overcome this resistance mechanism by inducing MMR-independent cell killing selectively in MGMT-silenced tumors. These agents deposit a dynamic DNA lesion that can be reversed by MGMT but slowly evolves into an interstrand cross-link in MGMT-deficient settings, resulting in MMR-independent cell death with low toxicity in vitro and in vivo. This discovery may lead to new treatments for gliomas and may represent a new paradigm for designing chemotherapeutics that exploit specific DNA repair defects.

Genetic instability is a hallmark of cancer and typically arises from mutations in key DNA damage repair and/or reversal proteins [collectively referred to here as the DNA damage response (DDR)] (*1*). Intrinsic DDR defects can be exploited with DDR inhibitors through the concept of synthetic lethality, which is defined as a loss of viability resulting from the disruption of two genes or pathways that, if disrupted individually, are nonlethal (*2*, *3*). Notable examples of synthetic lethal interactions between tumor-associated DDR defects and DDR inhibitors include (i) homologous recombination (HR)–defective tumors and inhibitors of polymerase θ (Pol θ) (*4*); (ii) ataxia-telangiectasia mutated (ATM)–mutant tumors and ataxia telangiectasia and Rad3-related (ATR) inhibitors (*5*); and (iii) mismatch repair (MMR)–deficient tumors and inactivation of Werner syndrome helicase (WRN) (*6*). HR-defective tumors also can be selectively targeted by inhibitors of poly(ADP-ribose) polymerase (PARP) (*2*, *3*). However, the mechanistic basis for this toxicity appears to depend on trapping of PARP at sites of DNA damage, which likely induces more complex damage that requires HR for resolution (*7*). In each of these examples, selective tumor cell killing by means of the DDR protein inhibitor relies on either the induction or persistence of DNA damage or aberrant DNA structures.

These findings suggested to us that genotoxins could be tailored to exploit differential sensitivities that arise from specific tumor-associated DDR defects. This approach avoids the need to engage DDR proteins directly, circumventing resistance mechanisms that arise from mutations in the ligand binding site (*8*) while minimizing off-target effects in healthy, DDR-proficient cells. To achieve this, we conceived a mechanistic strategy in which a single agent modifies DNA through two successive chemical steps (Fig. 1A). The first chemical reaction is designed to generate a primary DNA lesion that is rapidly removed by healthy, DDR-proficient cells. The second chemical reaction is engineered to slowly transform the primary modification into a more toxic secondary lesion. We anticipated that if the rate of primary lesion repair were sufficiently rapid in healthy cells, the secondary lesion would accumulate only in the DNA of DDR-deficient tumor cells. This two-step pathway would overcome resistance mechanisms that mitigate the toxicity of the primary lesion, which have been implicated in resistance to various chemotherapies, including anthracyclines [impairment of nucleoside excision repair (NER)] (*9*), topoisomerase inhibitors (loss of nonhomologous end joining repair) (*10*), and antimetabolite (*11*) and platinum (*12*) agents (mutations in MMR). Furthermore, delivering these bespoke lesions by using established chemotherapy scaffolds could facilitate rapid translation into the clinic, owing to decades of use in cancer patients. We examine this strategy in the context of glioblastoma (GBM).

GBM is the most common and devastating form of brain cancer, with a 5-year survival rate of ~5%. Approximately half of GBMs lack the DDR protein O^6-methylguanine methyltransferase (MGMT) through promoter hypermethylation. MGMT silencing occurs at an even greater frequency in grade II and III gliomas (more than 70% of cases), and these tumors are also largely incurable (*13*). Mechanistically, MGMT removes O^6-alkylguanine adducts by transferring the alkyl adduct to an active site cysteine through an S_N2 displacement. Patients with MGMT-deficient tumors (referred to here as MGMT– tumors) are treated with temozolomide (TMZ; **1a**), a prodrug that converts under physiological conditions to the potent methylating agent methyl diazonium (**1c**) through the intermediacy of 3-methyl-(triazen-1-yl)-imidazole-4-carboxamide (MTIC; **1b**) (Fig. 1B). N7-Methylguanine and N3-methyladenine are the major sites of methylation (70 and 9%, respectively) but are readily resolved by the base excision repair (BER) pathway. By contrast, although O^6-methylguanine (O^6MeG; **3**) adducts derived from TMZ (**1a**) only compose ~5% of alkylation products, these lesions persist in the genome of MGMT– tumors [but are readily reversed by healthy (MGMT+) cells] (Fig. 1C). O^6MeG (**3**) lesions are thought to induce formation of DNA double-strand breaks (DSBs) and tumor cell death through an MMR-dependent mechanism (*14*). MGMT status is a predictive biomarker for initial response to TMZ (**1a**) in GBM, with a significant overall survival benefit in the up-front setting for patients with these cancers (*15*). However, it is now well established (*16*) that acquired clinical resistance to TMZ (**1a**) through MMR mutations abrogates its efficacy, leading to recurrent GBM and death in nearly all patients. Frequently, TMZ (**1a**) is used as adjuvant therapy for grade III and high-risk grade II gliomas; however, it remains noncurative, with recurrences typically occurring over 2 to 10 years. In ~80% of patients, recurrences coincide with transformation to higher-grade tumors resistant to TMZ (**1a**), associated in ~50% of cases with acquisition of a distinct hypermutation signature secondary to MMR deficiency, resulting in reduced survival (*17*, *18*).

We hypothesized that we could deploy the strategy outlined in Fig. 1A to develop agents that overcome the resistance associated with MMR loss while maintaining TMZ's (**1a**) selectivity for MGMT-silenced tumors. These agents would deposit a primary lesion susceptible to S_N2-mediated removal by MGMT that could undergo a further chemical transformation to a secondary lesion capable of killing MGMT-deficient tumor cells in an MMR-independent manner. To maintain the therapeutic index (TI) between MGMT– tumor cells and MGMT+ healthy cells, the primary lesion must undergo MGMT-mediated repair faster than it undergoes transformation to the secondary lesion. With these considerations in mind, we developed an agent capable of depositing a 2-fluoroethyl lesion at O^6-guanine. We expected that O^6FEtG (**5**) would evolve slowly to the (N1,O^6Et)G intermediate **6**, which could then undergo ring-opening by N3 of the complementary cytidine base (**7**) to form the G(N1)–C(N3) interstrand cross-link (ICL) **8** (Fig. 1E). This hypothesis is based on the

[1]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [2]Department of Therapeutic Radiology, Yale School of Medicine, New Haven, CT 06520, USA. [3]Department of Pathology, Yale School of Medicine, New Haven, CT 06520, USA. [4]Department of Pharmacology, Yale School of Medicine, New Haven, CT 06520, USA.
*Corresponding author. Email: seth.herzon@yale.edu (S.B.H.); ranjit.bindra@yale.edu (R.S.B.)
†These authors contributed equally to this work.

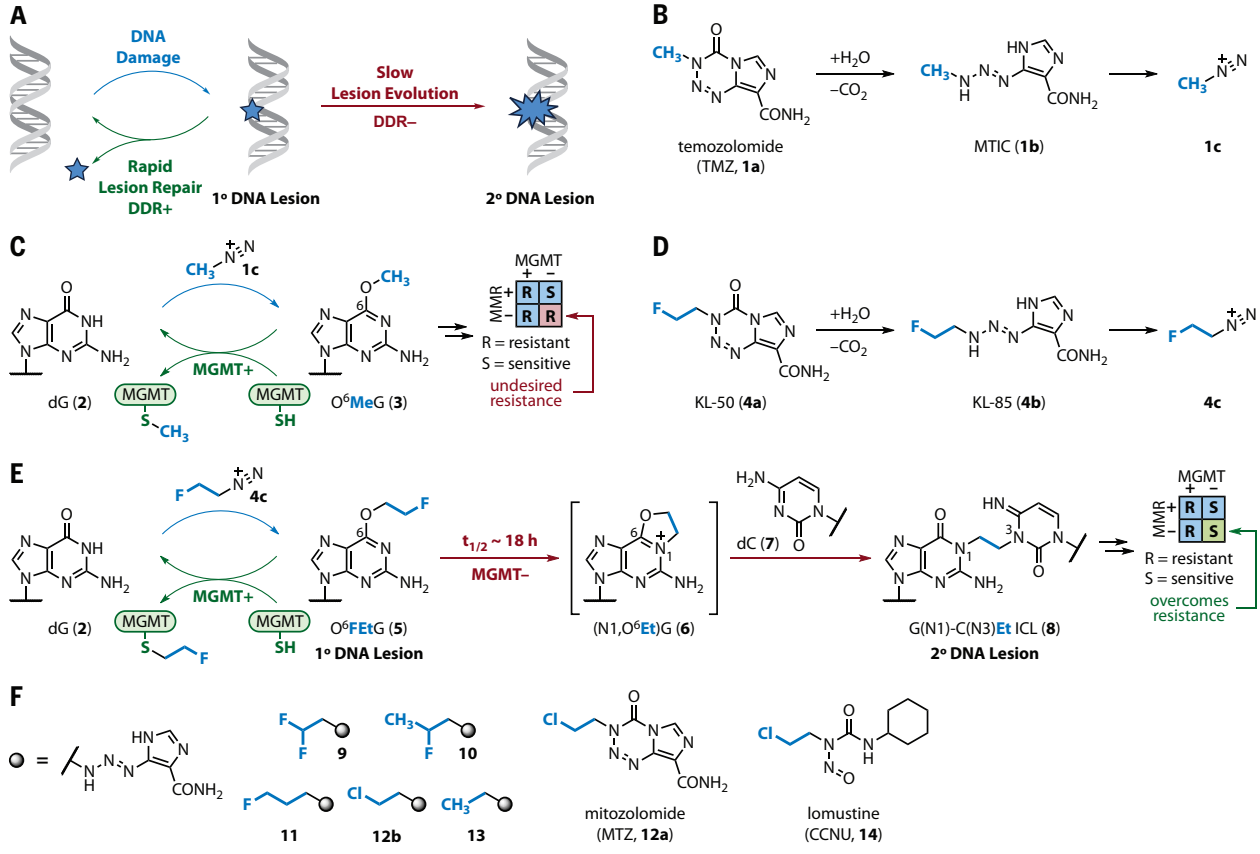


Fig. 1. Overview of mechanistic strategy and structures of agents used in this study. (**A**) Underlying mechanistic design. Systemic administration of a bifunctional agent is envisioned to form a primary lesion that is rapidly resolved by healthy (DDR+) but not DDR-deficient (DDR–) cells. The persistence of the primary lesion allows it to evolve slowly to a more toxic secondary lesion. (**B**) Under physiological conditions, TMZ (**1a**) converts to MTIC (**1b**), which in turn decomposes to methyl diazonium (**1c**). (**C**) O^6-Methylguanine (**3**) is the most clinical significant alkyl nucleobase generated by methyl diazonium (**1c**) and is rapidly reverted to dG (**2**) by MGMT. (**D**) Conversion of KL-50 (**4a**) to KL-85 (**4b**) in aqueous solution followed by decomposition to 2-fluoroethyl diazonium ion (**4c**). (**E**) 2-fluoroethyl diazonium ion (**4c**), derived from KL-50 (**4a**), is proposed to lead to the formation of G(N1)–C(N3)Et ICL **8** specifically in MGMT– cells. This mechanism is independent of MMR status. (**F**) Structures of the triazenes **9** to **13**, mitozolomide **12a**, and lomustine (CCNU; **14**). Syntheses are available in the supplementary materials.

finding that the corresponding nucleoside, O^6-(2-fluoroethyl)guanosine (**S1**), undergoes a hydrolytic rearrangement to N1-(2-hydroxyethyl)guanosine (**S3**) with a half-life of 18.5 hours (37°C, pH 7.4) (fig. S1A) (*19*) and the recent determination that oligonucleotides containing a single photocaged O^6-(2-chloroethyl)guanosine residue can undergo G–C cross-linking through a parallel mechanistic pathway (*20*). Although the rates of reversal of O^6FEtG (**5**) by MGMT are unknown, MGMT reacts rapidly with methylated calf thymus DNA (second-order rate constant = 1×10^9 M^{-1} min^{-1}) (*21*); and oligonucleotides that contain O^6-methyl-, O^6-benzyl-, and O^6-heteroarylmethylguanine residues possess nanomolar median inhibitory concentrations (IC_{50}) against MGMT in vitro (*22*). Additionally, MGMT reverses a wide range of O^6-alkylguanine substrates (*23*). These data suggested that MGMT-proficient cells should repair the O^6FEtG lesion (**5**) before it transforms into ICL **8**.

Results

Compound synthesis and in vitro evaluation

We synthesized the imidazotetrazine **4a** (also known as KL-50) and the triazene **4b** (also known as KL-85) as vehicles to deliver 2-fluoroethyl diazonium (**4c**) as well as a series of related agents to probe structure-activity relationships in tissue culture (Fig. 1, D and F). The 2-fluoropropyl- and 3-fluoropropyl-triazenes (**10** and **11**, respectively) were prepared through the diazotization of 4-aminoimidazole-5-carboxamide, followed by the addition of the respective amine (supplementary materials). All other triazenes were prepared according to literature procedures (supplementary materials). We evaluated the cytotoxicity of our compounds in short-term cell viability assays against four isogenic LN229 GBM cell lines engineered to be proficient or deficient in MGMT and/or MMR activity by using short hairpin RNAs (shRNAs) targeting MSH2 (referred to as MGMT+/–, MMR+/– cells hereafter) (fig. S2A). This approach allowed us to interrogate the relationship between MGMT and MMR status and compound activity.

The IC_{50} values of these agents are shown in Fig. 2A, and representative dose-response curves are shown in Fig. 2B (additional data are presented in fig. S2, B to G). KL-85 (**4b**) retained potency in MGMT–/MMR– cells (IC_{50} = 27.5 μM), whereas TMZ (**1a**) was essentially inactive (IC_{50} = 838 μM). Structure-activity studies were consistent with the mechanistic pathway shown in Fig. 1E. The 2,2-difluoroethyl triazene **9** and the 2-fluoropropyl triazene **10** possessed reduced potency in MGMT–/MMR– cells (fig. S2, B and C), which is in agreement with a reduced rate of displacement after introduction of an additional fluorine (*24*) or alkyl substituent. The 2-chloroethyl triazene **12b** was modestly potent but not as selective for MGMT– cell lines (fig. S2E), which likely results from faster, nonselective ICL formation arising from chloride displacement. The

3-fluoropropyl triazene **11** was essentially inactive in all four cell lines, potentially because of elimination of the diazonium ion (fig. S2D). As expected, the ethyl triazene **13** was inactive (fig. S2F). This compound is known to rapidly undergo elimination to ethylene gas after conversion to ethyl diazonium (*25*).

We prepared KL-50 (**4a**) through the diazotization of 4-aminoimidazole-5-carboxamide followed by the addition of (2-fluoroethyl) isocyanate (39% overall yield) (supplementary materials). The potency of **4a** mirrored that of **4b** in the four cell lines examined (Fig. 2B). To benchmark selectivity, we evaluated the experimental agent mitozolomide (MTZ; **12a**) and

A

IC_{50} (μM) of TMZ derivatives against MGMT+/–, MMR+/– LN229 cell lines.

Compound	Lesion Substituent	MGMT+ MMR+	MGMT+ MMR–	MGMT– MMR+	MGMT– MMR–	MGMT TI [a]	MMR RI [b]
TMZ (1a)	CH_3	742 ± 468	1005 ± 107	8.3 ± 3.5	838 ± 100	89	101
KL-50 (4a)	F	587 ± 166	504 ± 55	24.7 ± 4.1	22.0 ± 6.9	24	0.9
KL-85 (4b)	F	398 ± 97	481 ± 73	25.7 ± 4.6	27.5 ± 6.0	15	1.1
9	F, F	306 ± 88	410 ± 191	141 ± 47	982 ± 240	2.2	7.0
10	CH_3, F	660 ± 115	930 ± 183	334 ± 31	1004 ± 141	2.0	3.0
11	F	>1000	>1000	975 ± 63	>1000	---	---
MTZ (12a)	Cl	61.9 ± 4.7	50.8 ± 4.2	9.0 ± 0.4	6.9 ± 0.7	6.9	0.8
12b	Cl	59.8 ± 8.7	92.4 ± 11.5	12.4 ± 1.0	15.4 ± 1.8	4.8	1.2
13	CH_3	>1000	>1000	>1000	>1000	---	---
CCNU (14)	Cl	21.6 ± 3.2	26.8 ± 7.3	6.3 ± 0.3	4.2 ± 0.2	3.4	0.7

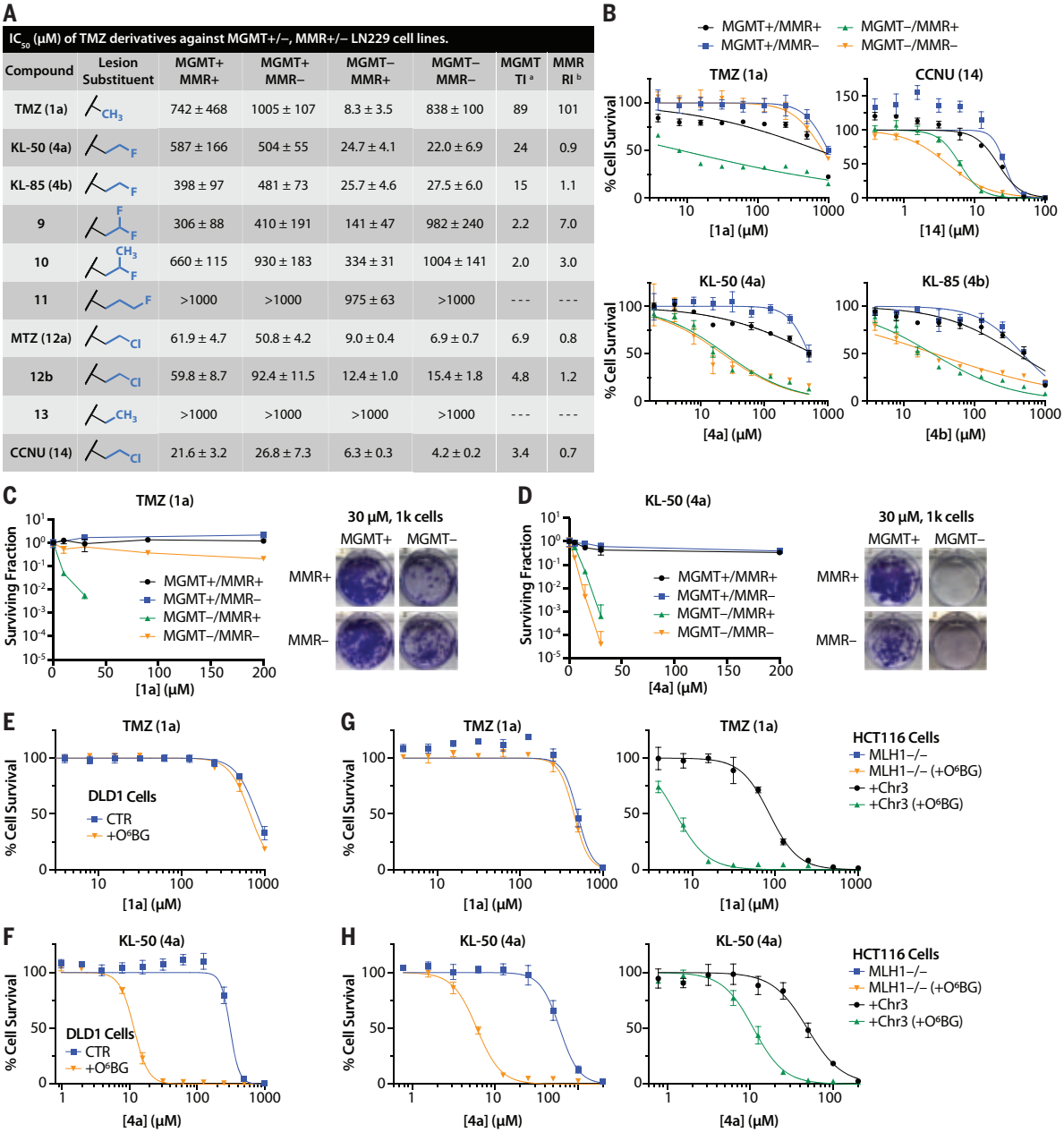


Fig. 2. KL-50 (4a) displays previously unseen MGMT-dependent, MMR-independent cytotoxicity in multiple isogenic cell models. (**A**) IC_{50} values derived from short-term viability assays in LN229 MGMT+/–, MMR+/– cells treated with TMZ (**1a**) derivatives. [a]MGMT TI = IC_{50} (MGMT+/MMR+) divided by IC_{50} (MGMT–/MMR+). [b]MMR resistance index (RI) = IC_{50} (MGMT–/MMR–) divided by IC_{50} (MGMT–/MMR+). (**B**) Short-term viability assay curves for TMZ (**1a**), CCNU (**14**), KL-85 (**4b**), and KL-50 (**4a**) in LN229 MGMT+/–, MMR+/– cells. (**C**) Clonogenic survival curves for TMZ (**1a**) in LN229 MGMT+/–, MMR+/– cells, with representative images of wells containing 1000 plated cells treated with 30 μM TMZ (**1a**). (**D**) Clonogenic survival curves for KL-50 (**4a**) in LN229 MGMT+/–, MMR+/– cells, with representative images of wells containing 1000 plated cells treated with 30 μM KL-50 (**4a**). (**E**) Short-term viability assay curves for TMZ (**1a**) in DLD1 MSH6-deficient cells pretreated with 0.01% dimethyl sulfoxide (DMSO) control (CTR) or 10 μM O^6BG (+O^6BG) for 1 hour before TMZ (**1a**) addition to deplete MGMT. (**F**) Short-term viability assay curves for KL-50 (**4a**) in DLD1 MSH6-deficient cells pretreated with 0.01% DMSO control (CTR) or 10 μM O^6BG (+O^6BG) for 1 hour before KL-50 (**4a**) addition. (**G**) Short-term viability assay curves for TMZ (**1a**) in HCT116 MLH1–/– cells or HCT116 cells complemented with chromosome 3 carrying wild-type MLH1 (+Chr3) pretreated with 0.01% DMSO control or 10 μM O^6BG (+O^6BG) for 1 hour before TMZ (**1a**) addition. (**H**) Short-term viability assay curves for KL-50 (**4a**) in HCT116 MLH1–/– cells or HCT116 cells complemented with chromosome 3 carrying wild-type MLH1 (+Chr3) pretreated with 0.01% DMSO control or 10 μM O^6BG (+O^6BG) for 1 hour before KL-50 (**4a**) addition. For (B) to (H), points indicate the mean, and error bars indicate SD; $n \geq 3$ technical replicates.

the clinical nitrosourea lomustine (also known as CCNU; **14**), which have been studied with hopes of addressing TMZ (**1a**) resistance (*26*, *27*). However, these 2-chloroethylating agents were only approximately four- to sevenfold selective for MGMT-deficient cells (fig. S2G and Fig. 2B, respectively), as opposed to the ~25-fold selectivity seen with the 2-fluoroethylating agent KL-50 (**4a**).

We validated the antitumor activity of KL-50 (**4a**) in clonogenic survival assays (CSAs) and additional cell lines in vitro. TMZ (**1a**) possessed negligible activity in MGMT+ LN229 cells, irrespective of MMR status, and induced robust tumor cell killing in MGMT–, MMR+ cells that was abolished in isogenic cells lacking MMR (Fig. 2C). Lomustine (**14**) was effective in MMR– cells but was cytotoxic to MGMT+ cells (fig. S2H). By contrast, KL-50 (**4a**) demonstrated robust antitumor activity in MGMT– cells, independent of MMR status, with negligable toxicity in MGMT+ cells at doses of up to at least 200 μM (Fig. 2D). We observed a similar pattern of activities in several distinct cell lines across different tumor types with intrinsic or induced loss of MGMT and/or MMR activity. For example, TMZ (**1a**) was inactive in DLD1 cells, which possess MGMT but lack functional MMR (MSH6–), with or without induced depletion of MGMT by using O^6-benzylguanine (O^6BG) (Fig. 2E) (*28*). By contrast, KL-50 (**4a**) was efficacious in these cells, but only after O^6BG-induced MGMT depletion (Fig. 2F). TMZ (**1a**) was inactive in HCT116 colorectal cancer cells, which lack the MMR protein MLH1, regardless of MGMT levels (Fig. 2G). Restoration of MMR activity by means of complementation with chromosome 3 containing MLH1 resulted in the expected enhanced sensitivity to TMZ (**1a**), which was further potentiated by MGMT depletion (Fig. 2G). By contrast, KL-50 (**4a**) induced selective tumor cell killing specifically in the setting of O^6BG-induced MGMT suppression in both MLH1-deficient and MLH1-complemented cells (Fig. 2H). The MMR status and O^6BG-induced loss of MGMT expression was confirmed with Western blot analysis for these cell lines (fig. S2, I and J). We also confirmed the activity of KL-50 (**4a**) in MGMT– LN229 cells engineered to lack expression of other key MMR proteins, including MSH6, MLH1, PMS2, and MSH3 (fig. S3). The inability of MSH3 depletion to induce TMZ (**1a**) resistance is an anticipated result, given that the primary role of the MSH2-MSH3 heterodimer is recognition of insertion-deletion loops, whereas single base pair mismatches induced by TMZ are processed by the MSH2-MSH6 heterodimer and subsequent downstream MMR proteins (*29*, *30*). Last, we compared the cytotoxicity of KL-50 (**4a**) and TMZ (**1a**) in normal human fibroblast cells and observed no increase in toxicity with KL-50 (**4a**) (fig. S2K). These data suggest that KL-50 (**4a**) is a first-in-class molecule that overcomes MMR mutation–induced resistance while retaining selectivity for tumor cells that lack MGMT.

Mechanism of action of KL-50 (4a)

We used a well-established comet assay adapted for ICL detection to determine whether ICLs were formed in MGMT– cells treated with KL-50 (**4a**) (Fig. 3, A and B). In this assay, cells were sequentially exposed to genotoxins and ionizing radiation (IR) and then analyzed by means of single-cell alkaline gel electrophoresis. Attenuation of the IR–induced comet tail is indicative of ICL formation. As expected, in the absence of IR, TMZ (**1a**) (200 μM) and KL-50 (**4a**) (200 μM) both induced tailing in MGMT–/MMR+ cells, whereas mitomycin C (MMC) (0.1 or 50 μM) did not (*31*). Exposure to 50 μM MMC for 2 hours abolished the IR-induced comet tail, whereas exposure to 0.1 μM MMC [chosen to be ~10-fold greater than the IC_{50} for this drug, which is comparable with 200 μM KL-50 (**4a**) or TMZ (**1a**)] for 24 hours caused a partial reduction in the IR-induced comet tail. TMZ (**1a**) (200 μM) did not reduce DNA migration after IR, which is in agreement with its known function as a monoalkylation agent with no known cross-linking activity. By contrast, KL-50 (**4a**) (200 μM) reduced the percent of DNA in the tail after IR to levels similar to those seen for 0.1 μM MMC. We observed a similar pattern of comet tail migration for MMC and KL-50 (**4a**) in MGMT–/MMR– cells, which supports an MMR-independent cross-linking mechanism. Comparable results were observed in MGMT–/MMR+ cells treated with KL-85 (**4b**) (fig. S4, A and B).

We carried out this assay at varying time points (2 to 24 hours) to assess the rates of ICL formation in MGMT–/MMR– cells treated with KL-50 (**4a**), MTZ (**12a**), or TMZ (**1a**) (Fig. 3, C and D). The chloroethylating agent MTZ (**12a**) reduced DNA mobility within 2 hours, which is consistent with the cell line selectivities above and literature reports that this agent rapidly forms ICLs (*32*) through chloride displacement from sites of 2-chloroethylation other than O^6G. TMZ (**1a**) did not induce a statistically significant decrease in DNA migration within 24 hours. However, we observed a time-dependent decrease in DNA mobility in cells treated with KL-50 (**4a**), with the largest difference observed between 8 and 24 hours; this is consistent with the reported half-life of O^6-(2-fluoroethyl)guanosine (**S1**) (18.5 hours) (fig. S1A) (*19*) and that of O^6FEtG (**6**) (18.5 hours). In the unirradiated samples, KL-50 (**4a**), MTZ (**12a**), and TMZ (**1a**) all induced maximal damage at 2 hours; this damage decreased over time, which is consistent with progressive DNA repair (fig. S4, C and D). Analysis of genomic DNA isolated from LN229 MGMT–/MMR+ cells treated with KL-50 (**4a**, 200 μM) or KL-85 (**4b**, 200 μM) by denaturing gel electrophoresis (*33*) demonstrated the presence of cross-linked DNA (Fig. 3E). TMZ (**1a**) and MTIC (**1b**) showed no evidence of ICL induction. Similarly, linearized pUC19 plasmid DNA treated with KL-50 (**4a**) (100 μM) also possessed ICLs, with delayed rates of formation relative to those of **12b** (Fig. 3F). Collectively, these data support a mechanism of action for KL-50 (**4a**) that involves the slow generation of DNA ICLs in the absence of MGMT.

To probe for alternative mechanisms of action, we conducted experiments to assess for involvement of nucleotide excision repair (NER) or BER in lesion processing, generation of reactive oxygen species (ROS), and DNA duplex destabilization. Short-term cell viability assays in isogenic mouse embryonic fibroblasts (MEFs) proficient or deficient in XPA, a common shared NER factor, revealed no differential sensitivity, either with or without O^6BG-induced MGMT depletion (fig. S5A). N7-MeG lesions induced by TMZ (**1a**) are prone to spontaneous depurination, apurinic (AP) site formation, and single-strand breaks (SSBs), which are all known BER substrates. To probe for potential differential induction of BER substrates by KL-50 (**4a**) compared with TMZ (**1a**), we performed in vitro supercoiled plasmid DNA assays to measure the formation of AP sites (*34*). We observed similar levels of spontaneous and enzyme-catalyzed SSBs from AP sites after treatment with KL-50 (**4a**) and TMZ (**1a**), suggesting comparable levels of depurination (fig. S5B). Cotreatment with increasing concentrations of the ROS scavenger *N*-acetylcysteine (NAC) did not rescue cell viability (fig. S5C). Melting point analysis did not reveal any notable differences in DNA stability resulting from fluoroethylation compared with methylation (fig. S5D). These data suggest that NER status, AP site induction, ROS, and altered DNA stability are peripheral or noncontributory to the effectiveness of KL-50 (**4a**).

We characterized the profile of DDR activation across our four isogenic cell lines after treatment with KL-50 (**4a**) or TMZ (**1a**). Our prior finding that the ATR-CHK1 signaling axis is activated in response to TMZ (**1a**)–induced replication stress in MGMT-deficient cells (*35*) prompted us to analyze the phosphorylation status of CHK1 and CHK2 in LN229 MGMT+/– and MMR+/– cells. KL-50 (**4a**) induced CHK1 and CHK2 phosphorylation in MGMT– cells regardless of MMR status, whereas TMZ (**1a**) only induced phospho-CHK1 and -CHK2 in MGMT–/MMR+ cells (fig. S6A). We analyzed foci formation of the DDR factors phospho-SER139-H2AX (γH2AX), p53 binding protein 1 (53BP1), and phospho-SER33-RPA2 (pRPA) over the period of 2 to 48 hours (Fig. 4, A to D, and fig. S6, B and C) (*36*). KL-50 (**4a**) induced a maximal foci response at 48 hours, specifically in MGMT– cells and irrespective of MMR status

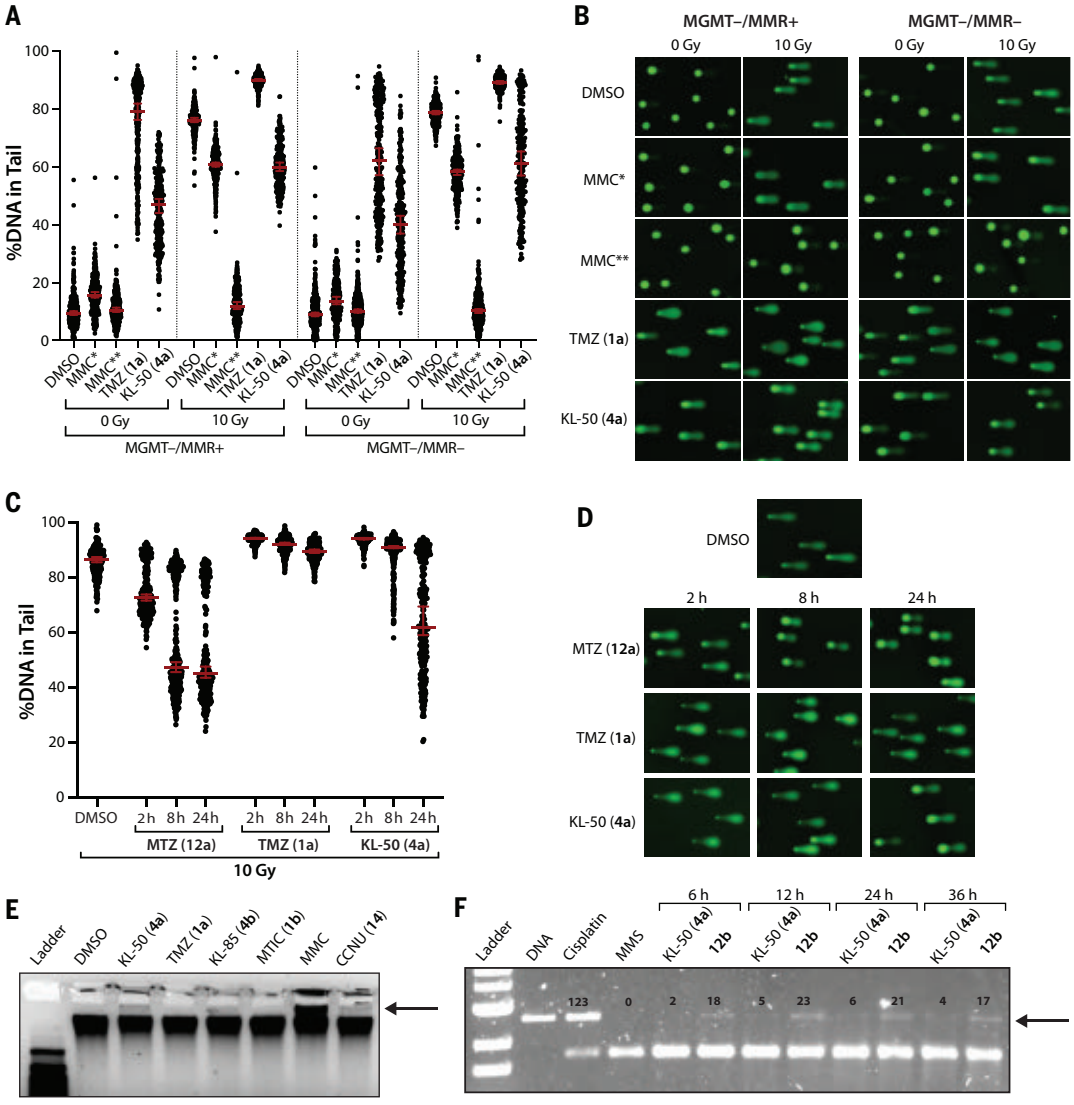


Fig. 3. Unrepaired primary KL-50 (4a) lesions convert to DNA ICLs in the absence of MGMT. (**A**) Scatter dot plots of the percent DNA in tail upon single-cell alkaline gel electrophoresis performed on LN229 MGMT−/MMR+ and MGMT−/MMR− cells treated with 0.2% DMSO control, 200 μM TMZ (**1a**), 200 μM KL-50 (**4a**), or 0.1 μM MMC (MMC*) for 24 hours or with 50 μM MMC (MMC**) for 2 hours. After cell lysis, comet slides were irradiated with 0 or 10 grays (Gy) before alkaline electrophoresis. Lines indicate median; error bars indicate 95% confidence interval (CI); $n \geq 160$ comets per condition. (**B**) Representative comet images from (A). (**C**) Scatter dot plots of the percent DNA in tail upon single-cell alkaline gel electrophoresis performed on LN229 MGMT−/MMR− cells treated with 0.2% DMSO control, 200 μM MTZ (**12a**), 200 μM TMZ (**1a**), or 200 μM KL-50 (**4a**) for 2, 8, or 24 hours. After cell lysis, comet slides were irradiated with 10 Gy before alkaline electrophoresis. Lines indicate median; error bars indicate 95% CI; $n \geq 230$ comets per condition. Data from samples treated with 0 Gy are shown in fig. S4, C and D. (**D**) Representative comet images from (C). (**E**) Denaturing gel electrophoresis of genomic DNA isolated from LN229 MGMT−/MMR+ cells treated with 0.2% DMSO control, 200 μM KL-50 (**4a**), 200 μM TMZ (**1a**), 200 μM KL-85 (**4b**), or 200 μM MTIC (**1b**) for 24 hours or with 50 μM MMC or 200 μM CCNU (**14**) for 2 hours. (**F**) Denaturing gel electrophoresis of linearized 100 ng pUC19 plasmid DNA treated in vitro with 100 μM cisplatin (36 hours), 100 μM MMS (36 hours), 100 μM of KL-50 (**4a**), or **12b** for 6 to 36 hours. For (E) and (F), bands indicating cross-linked DNA are indicated with arrows. Quantification of bands in (F) is provided in fig. S4E.

(**4a**). TMZ (**1a**) induced a comparable response in MGMT− cells, but this was abolished in the absence of functional MMR, which is consistent with known MMR-silencing–based resistance (*16*). We observed a reduced level of foci formation in MGMT+/MMR+ cells that was absent in MGMT+/MMR− cells, suggesting an MMR-dependent DDR in these cells. However, these foci dissipate at later time points (72 to 96 hours) (fig. S6D) and are not associated with appreciable cellular toxicity (Fig. 2, C and D).

KL-50 (**4a**) induced increasing G2 arrest on progression from 24 to 48 hours in MGMT−/MMR+ cells as determined with simultaneous analysis of DNA content based on nuclear (Hoechst) staining in the foci studies above (Fig. 4E and fig. S7, A and B). As previously reported for other sources of DNA damage, KL-50 (**4a**) induced an attenuated G2 arrest in MGMT−/MMR− cells, which is consistent with a role of MMR in the G2-checkpoint (*37*). This effect in MGMT−/MMR− cells was absent after TMZ (**1a**) treatment. Both TMZ (**1a**) and KL-50 (**4a**) induced a moderate G2 arrest in MGMT+/MMR+ cells.

We quantified the levels of DDR foci across the individual cell cycle phases (fig. S8). KL-50

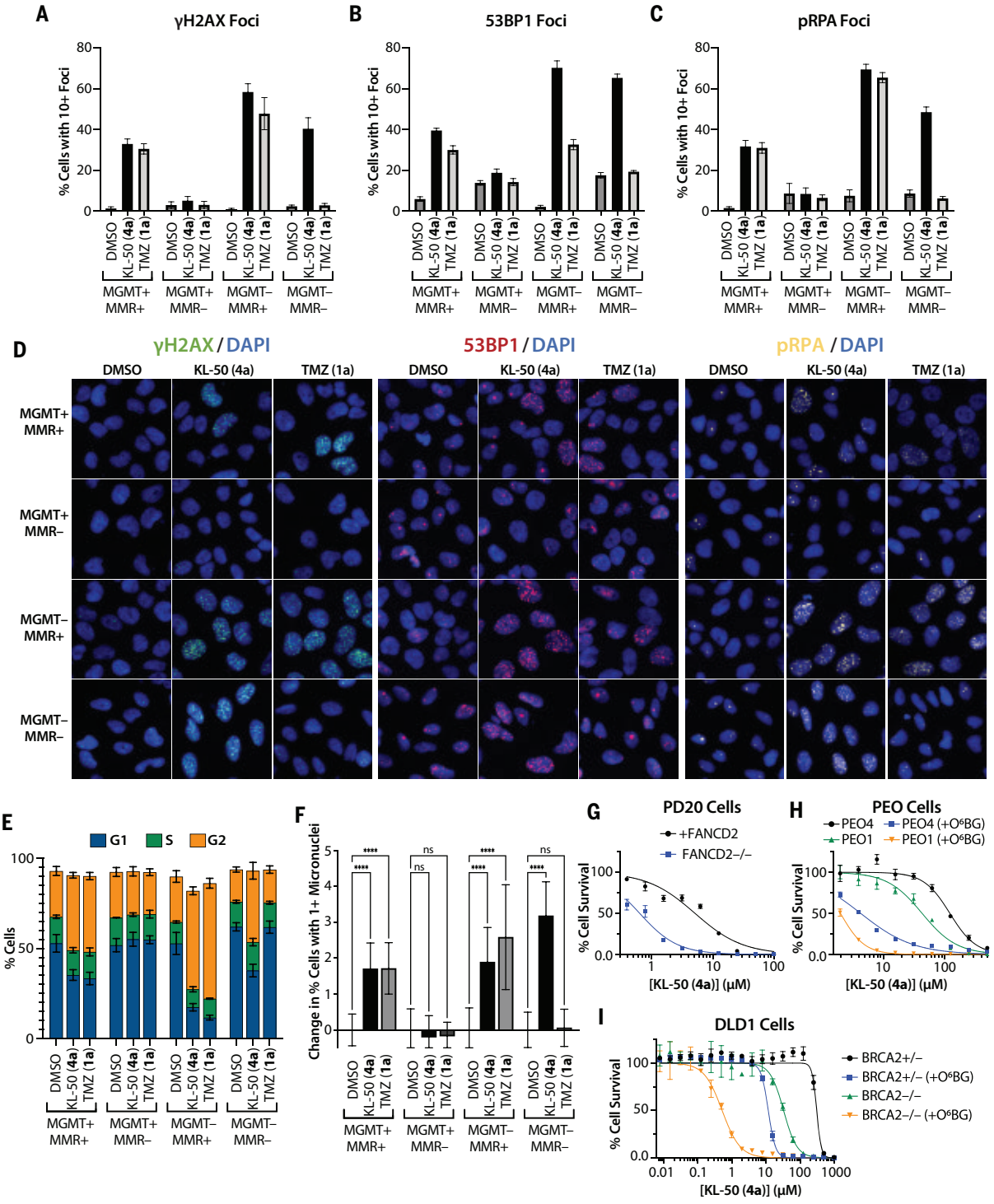


Fig. 4. KL-50 (4a) activates DDR pathways and cycle arrest in MGMT– cells, independent of MMR, and cells deficient in ICL or HR repair are sensitized to KL-50 (4a). (**A**) γH2AX, (**B**) 53BP1, and (**C**) pRPA foci formation quantified by percent of cells with ≥10 foci in LN229 MGMT+/–, MMR+/– cells treated with 0.1% DMSO control, 20 μM KL-50 (**4a**), or 20 μM TMZ (**1a**) for 48 hours. Columns indicate the mean; error bars indicate SD; $n \geq 5$ technical replicates. Additional time course data are presented in fig. S6, B to D. (**D**) Representative foci images of data in (A) to (C). (**E**) Percentage of cells in G_1, S, and G_2 cell cycle phases after treatment as in (A) to (C), measured by using integrated nuclear (Hoechst) staining intensity. Columns indicate the mean; error bars indicate SD; $n = 3$ independent analyses. Additional time course data, cell cycle controls, and representative histograms are presented in fig. S7. (**F**) Change in percent cells with ≥1 micronuclei from baseline (DMSO control) after treatment as in (A) to (C). Columns indicate the mean; error bars indicate SD; $n \geq 15$ technical replicates; ****$P < 0.0001$; ns, not significant. Additional validation is presented in fig. S9, A and B. (**G**) Short-term viability assay curves for KL-50 (**4a**) in PD20 cells, deficient in FANCD2 (FANCD2–/–) or complemented with FANCD2 (+FANCD2). (**H**) Short-term viability assay curves for KL-50 (**4a**) in PEO4 (BRCA2+) and PEO1 (BRCA2–/–) cells pretreated with 0.01% DMSO control or 10 μM O^6BG (+O^6BG) for 1 hour before KL-50 (**4a**) addition. (**I**) Short-term viability assay curves for KL-50 (**4a**) in DLD1 BRCA2+/– and BRCA2–/– cells pretreated with 0.01% DMSO control or 10 μM O^6BG (+O^6BG) for 1 hour before KL-50 (**4a**) addition. For (G) to (I), points indicate the mean, and error bars indicate SD; $n = 3$ technical replicates.

(**4a**) induced foci formation primarily in the S and G_2 phases of the cell cycle, which is consistent with replication blocking by ICLs (*38*). Foci increased in MGMT– G_1-phase cells at 48 hours, suggesting that a fraction of cells may progress through the S phase with unrepaired DNA damage. TMZ (**1a**) displayed a similar pattern of foci induction in the S and G_2 phases, with smaller increases in G_1-phase foci. We concurrently assessed micronuclei formation as an indication of replication stress-induced chromosomal instability (Fig. 4F). Consistent with previously reported data in MSH2–/– MEFs in which MMR loss suppressed micronuclei formation arising from replication-stress (*39*), MMR-deficient LN229 cells displayed suppression of micronuclei formation induced by KL-50 and TMZ in an MGMT+ background.

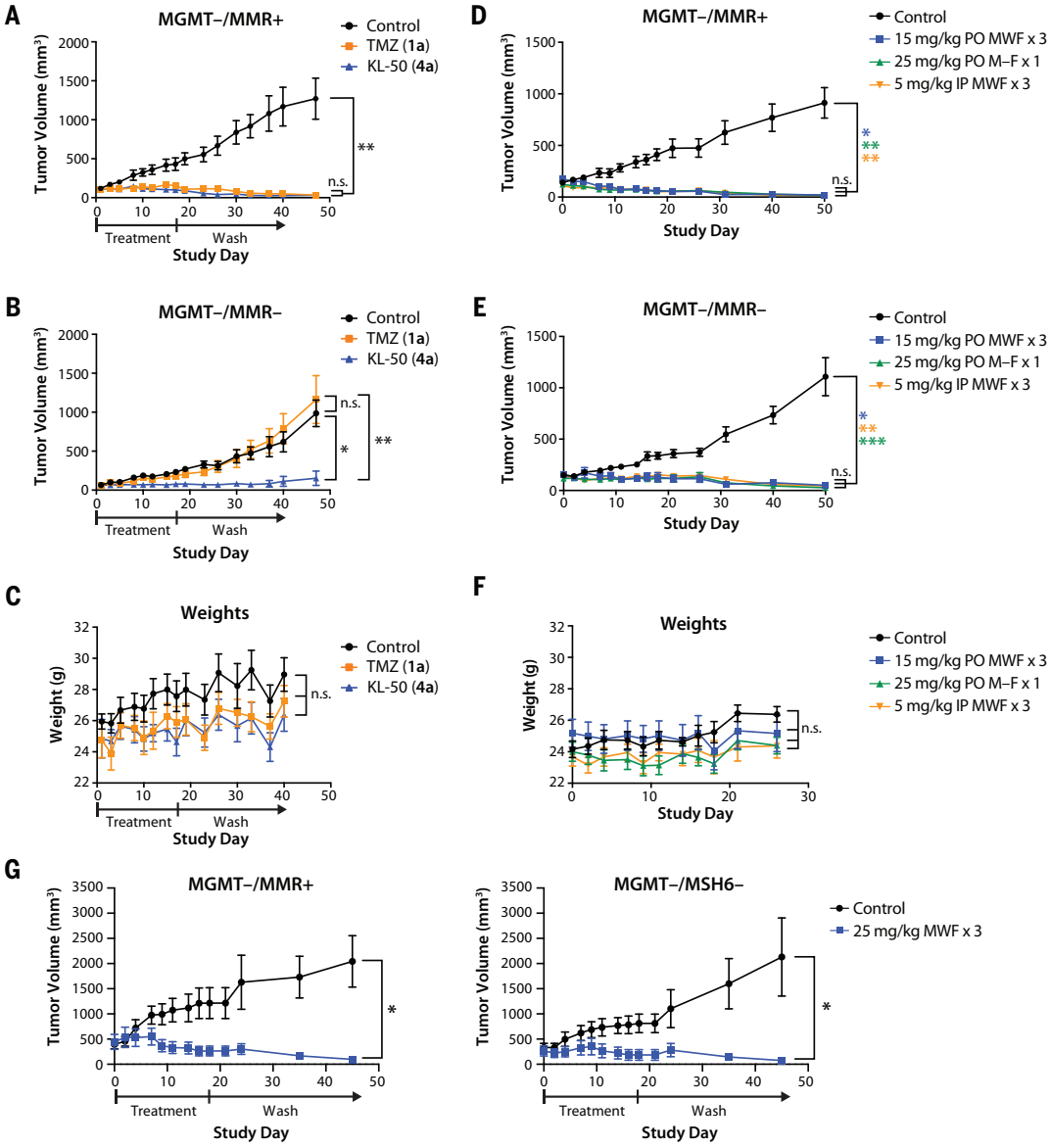


Fig. 5. KL-50 (4a) appears to be safe and efficacious in both MGMT–/MMR+ and MGMT–/MMR– flank tumors over a wide range of treatment regimens and conditions. (**A**) Xenograft LN229 MGMT–/MMR+ flank tumors treated with three weekly cycles of oral administration of 10% cyclodextrin control (n = 7 mice), TMZ (**1a**) (n = 7 mice, 5 mg/kg) or KL-50 (**4a**) (n = 6 mice, 5 mg/kg) on Monday, Wednesday, and Friday (individual spider plots in fig. S10A). (**B**) Xenograft LN229 MGMT–/MMR– flank tumors treated with three weekly cycles of oral administration of 10% cyclodextrin control (n = 6 mice), TMZ (**1a**) (n = 5 mice, 5 mg/kg), or KL-50 (**4a**) (n = 5 mice, 5 mg/kg) on Monday, Wednesday, and Friday (individual spider plots in fig. S10B). (**C**) Mean body weight of mice during LN229 flank tumor experiments in (A) and (B). (**D** to **F**) Xenograft (D) LN229 MGMT–/MMR+ and (E) LN229 MGMT–/MMR– flank tumors treated with oral administration of 10% cyclodextrin control (n = 7 mice), KL-50 (**4a**) (n = 6 mice, three cycles of 15 mg/kg on Monday, Wednesday, and Friday), KL-50 (**4a**) (n = 6 mice, 1 cycle of 25 mg/kg Monday through Friday), or intraperitoneal administration of KL-50 (**4a**) (n = 7 mice, three cycles of 5 mg/kg on Monday, Wednesday, and Friday) revealed equal efficacy with no observable increases in toxicity, (F) as measured with mice systemic weights (individual spider plots are provided in fig. S10C). (**G**) Xenograft LN229 MGMT–/MMR+ and LN229 MGMT–/MSH6– flank tumors with a larger average starting tumor size of ~400 mm^3 and ~350 mm^3, respectively, treated with three weekly cycles of oral administration of 10% cyclodextrin (n = 4 mice) or KL-50 (**4a**) (n = 4 mice, 25 mg/kg on Monday, Wednesday, and Friday). The study period was limited by control groups that had to be euthanized for exceeding the ethical maximum allowed tumor size, thus ending the study. In (A) to (G), points indicate the mean, and error bars indicate SEM; *P < 0.05; **P < 0.01; ***P < 0.001; n.s., not significant.

However, KL-50 (**4a**), but not TMZ (**1a**), was able to induce micronuclei in MGMT–/MMR– cells, which suggests that the ICLs formed in these cells are able to induce chromosomal instability even in the absence of MMR. These findings are in agreement with the differential toxicity profiles of KL-50 (**4a**) and TMZ (**1a**): KL-50 (**4a**) induces multiple successive markers of DNA damage and engagement of the DDR in MGMT– cells, independent of MMR status, whereas the effects of TMZ (**1a**) are similar to those of KL-50 in MGMT–/MMR+ cells but absent in MMR– cells. Coupled with the ICL kinetics data presented above, these time-course data support a slow rate of ICL induction in situ by KL-50 (**4a**).

These foci data suggest that KL-50 (**4a**) induces replication stress (as indicated by pRPA foci formation) and formation of DSBs (as indicated by γH2AX and 53BP1 foci), which are known to follow the formation of ICLs (*38*). Consistent with this, BRCA2- and FANCD2-deficient cells are hypersensitive to KL-50 (**4a**) (Fig. 4, G to I, and fig. S9, C to F). In two MGMT-proficient cell models, BRCA2 loss enhanced the toxicity of KL-50 (**4a**) after MGMT depletion through O^6BG (Fig. 4, H and I). We observed FANCD2 ubiquitination by KL-50 (**4a**) specifically in MGMT– cells, suggesting activation of the Fanconi anemia (FA) ICL repair pathway (fig. S9G). As previously reported (*40*), TMZ (**1a**) also induced FANCD2 ubiquitination, but only in MGMT–/MMR+ cells. This is an expected result given the propensity of O^6MeG (**3**) to induce stalled replication forks through the process of futile cycling (*41*).

KL-50 (4a) is active against TMZ (1a)–resistant tumors in vivo

We evaluated the activity of KL-50 (**4a**) and TMZ (**1a**) in vivo using mouse flank tumor models derived from the isogenic LN229 MGMT– cell lines. We treated MGMT–/MMR+ and MGMT–/MMR– flank tumors with KL-50 (**4a**) or TMZ (**1a**) (5 mg/kg orally Monday, Wednesday, and Friday for 3 weeks) as previously described for TMZ (**1a**) (*35*). As expected, TMZ (**1a**) suppressed tumor growth in the MGMT–/MMR+ tumors (Fig. 5A). KL-50 (**4a**) was statistically noninferior to TMZ (**1a**), despite a 17% lower molar dosage owing to its higher molecular weight. In the MGMT–/MMR– tumors, TMZ (**1a**) demonstrated no efficacy, whereas KL-50 (**4a**) potently suppressed tumor growth (Fig. 5B). KL-50 (**4a**) treatment resulted in no statistically significant changes in body weight compared with that of TMZ (**1a**) or control (Fig. 5C). KL-50 (**4a**) was effective and nontoxic across different dosing regimens (5, 15, and 25 mg/kg), treatment schedules (Monday, Wednesday, and Friday for 3 weeks, and Monday through Friday for 1 week), and routes of drug administration (oral or intraperitoneal) in mice bearing MGMT–/MMR+ and MGMT–/MMR– flank tumors (Fig. 5, D to F). KL-50 (**4a**) (25 mg/kg orally Monday, Wednesday, and Friday for 3 weeks) potently suppressed the growth of large (~350 to 400 mm^3) MGMT–/MMR+ and MGMT–/MSH6– tumors (Fig. 5G). KL-50 (**4a**) (25 mg/kg orally Monday through Friday for 1 week) was also effective in an orthotopic, intracranial LN229 MGMT–/MMR– model, whereas TMZ (**1a**) only transiently suppressed tumor growth (Fig. 6, A and B).

A focused maximum tolerated dose study revealed that KL-50 (**4a**) is well tolerated. Healthy mice were treated with escalating doses of KL-50 (**4a**) (0, 25, 50, 100, and 200 mg/kg orally for 1 dose) and monitored over time for changes in both weights and hematologic profiles. Mice in the higher dosage groups (100 or 200 mg/kg) experienced a greater than 10% weight loss after treatment, which regressed to baseline at the end of 1 week (Fig. 6C). Two of three mice in the 200 mg/kg treatment group became observably ill, warranting euthanasia, but no evidence of toxicity was observed in the remaining cohorts. Because the main

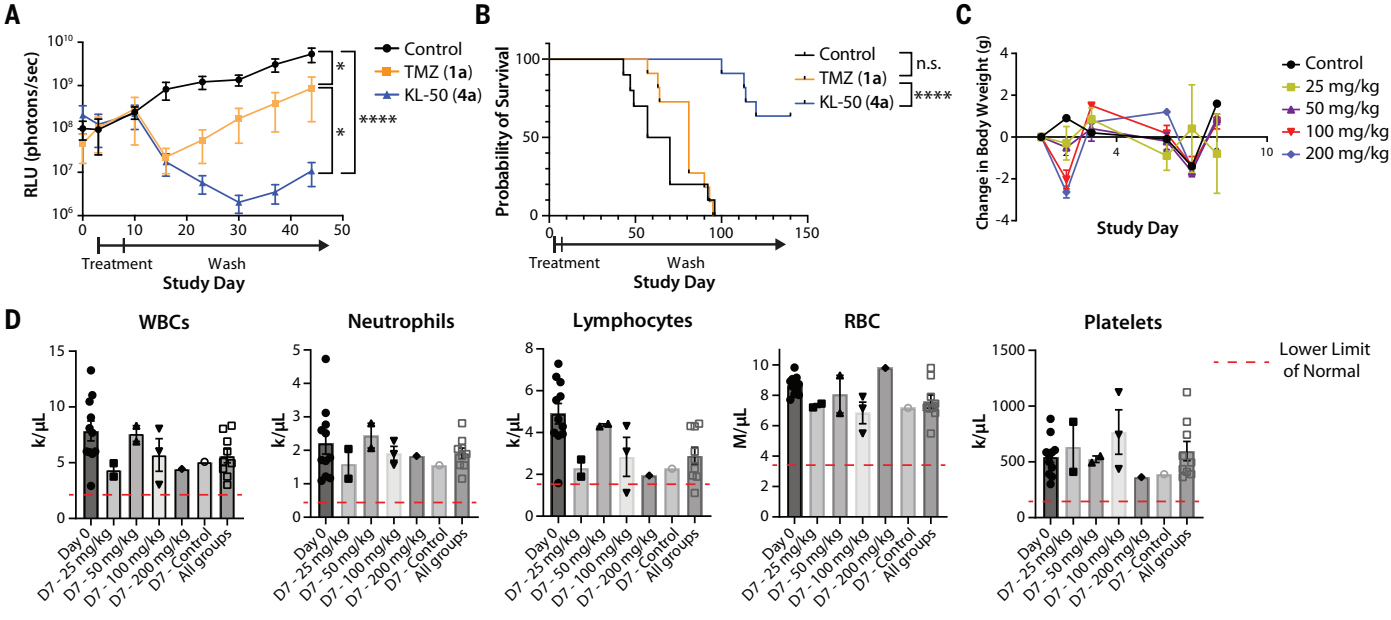


Fig. 6. KL-50 (4a) is efficacious in an LN229 MGMT–/MMR– intracranial model and is well tolerated with limited myelosuppression at supratherapeutic doses. (**A**) Mean tumor size as measured with bioluminescent imaging as relative light units (RLU; photons/s) with SEM of xenograft LN229 MGMT–/MMR– intracranial tumors treated with 5 consecutive days of oral administration of 10% cyclodextrin control (n = 10 mice), TMZ (**1a**) (n = 11 mice, 25 mg/kg), or KL-50 (**4a**) (n = 11 mice, 25 mg/kg) (individual spider plots are provided in fig. S10D). (**B**) Kaplan-Meier analysis of intracranial xenograft tumor-bearing mice in (A), treated with 5 consecutive days of oral administration of 10% cyclodextrin control (n = 10 mice), TMZ (**1a**) (n = 11 mice, 25 mg/kg), or KL-50 (**4a**) (n = 11 mice, 25 mg/kg) demonstrating a significant survival benefit for KL-50 (**4a**) compared with both control and TMZ (**1a**) groups. (**C**) Mean body weight change with SEM of mice during maximum tolerated dose experiment in non–tumor-bearing mice. (**D**) Complete blood counts for mice before treatment and 7 days after treatment with escalations of single-dose KL-50 (**4a**) delivered orally. White blood cells (WBC) lower limit of normal (LLN), 2.2 k/µL; neutrophil LLN, 0.42 k/µl; lymphocyte LLN, 1.7 k/µl; red blood cells (RBC) LLN, 3.47 M/µl; platelet LLN, 155 k/µl. *P < 0.05; ****P < 0.0001; n.s.; not significant.

dose-limiting systemic toxicity of TMZ (**1a**) is myelosuppression, we measured complete blood counts for all mice on day 0 before treatment and subsequently on day 7 after drug administration. Overall, there was a trend toward reduced neutrophil and lymphocyte counts with KL-50 (**4a**) treatment, although average blood counts were within normal physiological ranges (defined as values falling within 2 SD of the average for healthy mice) for all cohorts (Fig. 6D). Taken together, these data demonstrate in vivo efficacy, acceptable systemic tolerability, and central nervous system penetrance of KL-50 (**4a**).

Discussion

MMR mutation–induced resistance to alkylating agents has been a major barrier to treatment efficacy since the introduction of TMZ (**1a**) in the early 1990s. Chloroethyl alkylating agents, such as lomustine (**14**) and MTZ (**12a**), lack a TI owing to indiscriminate cross-linking activity through chloride displacement (*32*, *42*). Fluoride is not normally susceptible to bimolecular displacement [the homolytic bond dissociation energies of aliphatic C–F and C–Cl bonds are ~115 kcal/mol (481 kJ/mol) and 83.7 kcal/mol (350 kJ/mol), respectively]. However, the appropriate positioning of hydrogen bond donors or covalently attached nucleophiles can promote substitution (*43*). Thus, the intramolecular displacement pathway (**5**→**6**) appears to provide an essential acceleration in the formation of ICLs by KL-50 (**4a**) while maintaining sufficient lifetime for MGMT-mediated repair of the O^6FEtG (**5**) adduct. This hypothesis is consistent with the reported half-lives of O^6-(2-fluoroethyl)guanosine (**S1**) and O^6-(2-chloroethyl)guanosine (**S4**) ($t_{1/2}$ = ~18.5 hours and ~18 min) (fig. S1, A and B) (*19*, *44*) and the reported resistance of N7-(2-fluoroethyl)guanosine (**S5**) to bimolecular fluoride displacement (fig. S1C) (*45*). We hypothesize that the rate of MGMT reversal of O^6FEtG (**5**) is comparable with the corresponding ethyl and *n*-propyl adducts, which are processed at 25 to 50% the rate of O^6MeG (**3**) in vitro (*46*). It has been reported that the mean lethal dose of ICLs in HeLa cells is 230 (*47*), and TMZ (**1a**) has been demonstrated to yield 20,600 O^6MeG (**3**) adducts per cell at a dose of 20 μM (*48*). Assuming that a similar level of O^6FEtG (**5**) lesions are induced by KL-50 (**4a**) at the IC_{50} (~20 μM) in MGMT–/MMR– LN229 cells, the number of adducts required to convert to ICLs to generate the mean lethal dose is ~1 in 90, or ~1.1% cross-linking efficiency.

Beyond MGMT-silenced recurrent glioma, it is possible that there may be additional beneficial indications for selective targeting of cancer cells with KL-50 (**4a**). MGMT silencing has been reported in 40% of colorectal cancers and 25% of non–small cell lung cancer, lymphoma, and head and neck cancers (*49*). MGMT mRNA expression is also reduced in subsets of additional cancer types, including breast carcinoma, bladder cancer, and leukemia (*50*). MMR loss, as reported by microsatellite instability, is a well-established phenomenon in multiple cancer types and leads to resistance to various standard-of-care agents.

Our data also suggest that KL-50 (**4a**) will display a higher TI in tumors with MGMT deficiency and impaired ICL repair, including HR deficiency. The TI of KL-50 (**4a**) in the DLD1 isogenic model, as measured by the ratio of IC_{50} values in MGMT+/BRCA2+ cells compared with MGMT–/BRCA2– cells, was ~600-fold, which is larger than canonical cross-linking agents cisplatin (42-fold) or MMC (26-fold). A similar amplification of the TI was seen in the PEO1/4 model with KL-50 (**4a**) (62-fold) versus cisplatin (13-fold) or MMC (7-fold). HR-related gene mutations have been detected in a substantial number of tumors across multiple cancer types (17.4% in 21 cancer lineages), and methods have been developed to assess for tumor-associated HR deficiency (*51*).

Approximately half of patients with recurrent MGMT-methylated glioma acquire TMZ (**1a**) resistance through loss of MMR (*18*, *52*). KL-50 (**4a**) is designed to fill this therapeutic void. We expect KL-50 (**4a**) to be amenable to derivatization for improved drug pharmacokinetic properties on the basis of prior work with the imidazotetrazine scaffold (*53*), facilitating translation to clinical trials. Broadly, incorporating a consideration of the relative rates of DNA modification and DNA repair pathways into the design of therapeutics may lead to the development of additional selective chemotherapies.

REFERENCES AND NOTES

1. N. J. Curtin, *Nat. Rev. Cancer* **12**, 801–817 (2012).
2. H. E. Bryant *et al.*, *Nature* **434**, 913–917 (2005).
3. H. Farmer *et al.*, *Nature* **434**, 917–921 (2005).
4. D. Zatreanu *et al.*, *Nat. Commun.* **12**, 3636 (2021).
5. P. M. Reaper *et al.*, *Nat. Chem. Biol.* **7**, 428–430 (2011).
6. S. Lieb *et al.*, *eLife* **8**, e43333 (2019).
7. J. Murai *et al.*, *Cancer Res.* **72**, 5588–5599 (2012).
8. M. P. Dias, S. C. Moser, S. Ganesan, J. Jonkers, *Nat. Rev. Clin. Oncol.* **18**, 773–791 (2021).
9. M. A. Sabatino *et al.*, *Mol. Cancer* **9**, 259 (2010).
10. N. Adachi, S. So, H. Koyama, *J. Biol. Chem.* **279**, 37343–37348 (2004).
11. S. E. Fordham, E. C. Matheson, K. Scott, J. A. E. Irving, J. M. Allan, *Leukemia* **25**, 1046–1049 (2011).
12. A. Sawant, A. Kothandapani, A. Zhitkovich, R. W. Sobol, S. M. Patrick, *DNA Repair* **35**, 126–136 (2015).
13. E. H. Bell *et al.*, *JAMA Oncol.* **4**, 1405–1409 (2018).
14. G.-M. Li, *Cell Res.* **18**, 85–98 (2008).
15. M. E. Hegi *et al.*, *N. Engl. J. Med.* **352**, 997–1003 (2005).
16. K. Yoshimoto *et al.*, *Front. Oncol.* **2**, 186 (2012).
17. M. Touat *et al.*, *Nature* **580**, 517–523 (2020).
18. Y. Yu *et al.*, *Neuro-oncol.* **23**, 1872–1884 (2021).
19. W. P. Tong, M. C. Kirk, D. B. Ludlum, *Biochem. Pharmacol.* **32**, 2011–2015 (1983).
20. S. Hentschel, J. Alzeer, T. Angelov, O. D. Schärer, N. W. Luedtke, *Angew. Chem. Int. Ed.* **51**, 3466–3469 (2012).
21. C. L. Chan, Z. Wu, T. Ciardelli, A. Eastman, E. Bresnick, *Arch. Biochem. Biophys.* **300**, 193–200 (1993).
22. T. Shibata *et al.*, *Nucleic Acids Res.* **34**, 1884–1891 (2006).
23. A. E. Pegg, *Chem. Res. Toxicol.* **24**, 618–639 (2011).
24. H. Martinez, A. Rebeyrol, T. B. Nelms, W. R. Dolbier Jr., *J. Fluor. Chem.* **135**, 167–175 (2012).
25. Y. Wang, R. T. Wheelhouse, L. Zhao, D. A. F. Langnel, M. F. G. Stevens, *J. Chem. Soc., Perkin Trans. 1* **10**, 1669–1676 (1998).
26. J. A. Double, M. C. Bibby, *J. Natl. Cancer Inst.* **81**, 988–994 (1989).
27. L. J. Fairbairn, N. Chinnasamy, L. S. Lashford, D. Chinnasamy, J. A. Rafferty, *Cancer Gene Ther.* **7**, 233–239 (2000).
28. L. Liu, S. Markowitz, S. L. Gerson, *Cancer Res.* **56**, 5375–5379 (1996).
29. J. Jiricny, *Cold Spring Harb. Perspect. Biol.* **5**, a012633 (2013).
30. S. Ganesa, A. Sule, R. K. Sundaram, R. S. Bindra, *Sci. Rep.* **12**, 5827 (2022).
31. D. J. McKenna, M. Gallus, S. R. McKeown, C. S. Downes, V. J. McKelvey-Martin, *DNA Repair* **2**, 879–890 (2003).
32. T. P. Brent, P. E. Gonzaga, D. G. Smith, in *DNA Repair Mechanisms and Their Biological Implications in Mammalian Cells*, M. W. Lambert, J. Laval, Eds. (Springer, 1989), pp. 619–638.
33. N. Bossuet-Greif *et al.*, *mBio* **9**, e02393-17 (2018).
34. M. Xue, K. M. Wernke, S. B. Herzon, *Biochemistry* **59**, 892–900 (2020).
35. C. B. Jackson *et al.*, *Cancer Res.* **79**, 4331–4338 (2019).
36. K. Rothkamm *et al.*, *Environ. Mol. Mutagen.* **56**, 491–504 (2015).
37. V. O'Brien, R. Brown, *Carcinogenesis* **27**, 682–692 (2006).
38. D. Vare *et al.*, *DNA Repair* **11**, 976–985 (2012).
39. Y. Matsuno *et al.*, *Nat. Commun.* **10**, 3925 (2019).
40. C. C. Chen, T. Taniguchi, A. D'Andrea, *J. Mol. Med.* **85**, 497–509 (2007).
41. N. Mojas, M. Lopes, J. Jiricny, *Genes Dev.* **21**, 3342–3355 (2007).
42. N. Leelatian, C. S. Hong, R. S. Bindra, *Neurosurg. Clin. N. Am.* **32**, 171–180 (2021).
43. Y. Pan, *ACS Med. Chem. Lett.* **10**, 1016–1019 (2019).
44. S. Parker, M. C. Kirk, D. B. Ludlum, *Biochem. Biophys. Res. Commun.* **148**, 1124–1128 (1987).
45. D. B. Ludlum, W. P. Tong, *Biochem. Pharmacol.* **27**, 2391–2394 (1978).
46. K. Morimoto, M. E. Dolan, D. Scicchitano, A. E. Pegg, *Carcinogenesis* **6**, 1027–1031 (1985).
47. P. D. Lawley, D. H. Phillips, *Mutation Res.* **355**, 13 (1996).
48. B. Stratenwerth *et al.*, *Mol. Cancer Ther.* **20**, 1789–1799 (2021).
49. M. Esteller, S. R. Hamilton, P. C. Burger, S. B. Baylin, J. G. Herman, *Cancer Res.* **59**, 793–797 (1999).
50. A. Thomas *et al.*, *Cancer Res.* **77**, 823–826 (2017).
51. H. Davies *et al.*, *Nat. Med.* **23**, 517–525 (2017).
52. D. P. Cahill, P. J. Codd, T. T. Batchelor, W. T. Curry, D. N. Louis, *Clin. Neurosurg.* **55**, 165–171 (2008).
53. R. L. Svec *et al.*, *ACS Chem. Biol.* **13**, 3206–3216 (2018).

ACKNOWLEDGMENTS

We thank S. Ganesa for her help in generating MMR-deficient isogenic cell lines used in this work. We thank the Yale Center for Molecular Discovery for their assistance with high-throughput immunofluorescence microscopy and image analysis. **Funding:** We gratefully acknowledge financial support from the National Institutes of Health (R35-GM131913 to S.B.H, T32-GM007205 and F30-CA254158 to K.L, and 5R01CA215453-02 and U19CA264362 to R.S.B.), the Yale Cancer Center (Pilot grant to S.B.H. and R.S.B.), and the CureSearch for Children's Cancer (Catapult Award to R.S.B.). The Yale Center for Molecular Discovery is supported in part by an NCI Cancer Center Support Grant NIH P30 CA016359. Equipment and libraries were supported in part by the Program in Innovative Therapeutics for Connecticut's Health. **Author contributions:** S.B.H. and R.S.B. conceived and codirected the study. K.L., S.E.G., R.K.S., and E.D.H. carried out the experiments. All authors contributed to the design and analysis of the experiments and writing of the manuscript. **Competing interest:** S.B.H., R.S.B., and K.L. are cofounders and hold equity in Aztek Biosciences, which has exclusive rights to the findings herein. K.L., S.E.G., R.K.S., E.D.H., R.S.B., and S.B.H. are inventors on US Provisional Application 63/290,572 submitted by Yale University, which covers the mechanism of action and use of compound **4a** for the treatment of MGMT–/MMR– GBM. K.L., R.S.B., and S.B.H are inventors on US Provisional Applications 63/209,763, 63/212,410, 63/247,645, 63/290,622, 63/290,627, and 63/290,630, which cover compounds **4a**, **4b**, and related derivatives for the treatment of MGMT– tumors. **Data and materials availability:** Crystallographic data for KL-50 (**4a**) has been deposited in the Cambridge Crystallographic Data Centre as CCDC 2122008/2122009. All other data are included in the manuscript and supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim

to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS
science.org/doi/10.1126/science.abn7570

Materials and Methods
Figs. S1 to S12
Synthetic Procedures
Crystallographic Analysis of KL-50 (**4a**)
Catalog of Nuclear Magnetic Resonance Spectra

References (*54*–*68*)
MDAR Reproducibility Checklist

Submitted 17 December 2021; accepted 10 June 2022
10.1126/science.abn7570

NEURODEVELOPMENT

Analysis of somatic mutations in 131 human brains reveals aging-associated hypermutability

Taejeong Bae[1]†, Liana Fasching[2]†, Yifan Wang[1]†, Joo Heon Shin[3,4]†, Milovan Suvakov[1], Yeongjun Jang[1], Scott Norton[2], Caroline Dias[5,6]‡, Jessica Mariani[2], Alexandre Jourdon[2], Feinan Wu[2], Arijit Panda[1], Reenal Pattni[7], Yasmine Chahine[5,6], Rebecca Yeh[5,6], Rosalinda C. Roberts[8], Anita Huttner[9], Joel E. Kleinman[3,10], Thomas M. Hyde[3,4,10], Richard E. Straub[3], Christopher A. Walsh[5,6], Brain Somatic Mosaicism Network§, Alexander E. Urban[7], James F. Leckman[2], Daniel R. Weinberger[3,4,10,11,12]*, Flora M. Vaccarino[2,13]*, Alexej Abyzov[1]*

We analyzed 131 human brains (44 neurotypical, 19 with Tourette syndrome, 9 with schizophrenia, and 59 with autism) for somatic mutations after whole genome sequencing to a depth of more than 200×. Typically, brains had 20 to 60 detectable single-nucleotide mutations, but ~6% of brains harbored hundreds of somatic mutations. Hypermutability was associated with age and damaging mutations in genes implicated in cancers and, in some brains, reflected in vivo clonal expansions. Somatic duplications, likely arising during development, were found in ~5% of normal and diseased brains, reflecting background mutagenesis. Brains with autism were associated with mutations creating putative transcription factor binding motifs in enhancer-like regions in the developing brain. The top-ranked affected motifs corresponded to MEIS (myeloid ectopic viral integration site) transcription factors, suggesting a potential link between their involvement in gene regulation and autism.

Somatic mutations naturally occur in proliferative and postmitotic cells throughout human development and during aging, starting from the first cleavage of the zygote (*1*–*4*). It has yet to be determined how frequent somatic mutations are in the population and whether they are a contributing factor to the etiology of neuropsychiatric disorders (*5*). Mutations analyzed in bulk tissues are typically present in >1% of cells, often arise during early development, and may be enriched for variants exerting a stronger phenotypic effect. We investigated somatic mosaicism in normal human brains (n = 44) and in brains of those affected by one of three neuropsychiatric diseases: Tourette syndrome (TS; n = 19), schizophrenia (SCZ; n = 9), or autism spectrum disorder (ASD; n = 59).

[1]Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA. [2]Child Study Center, Yale University, New Haven, CT, USA. [3]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA. [4]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [5]Division of Genetics and Genomics and Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA. [6]Department of Pediatrics and Department of Neurology, Harvard Medical School, Boston, MA, USA. [7]Department of Psychiatry and Behavioral Sciences and Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [8]Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, USA. [9]Department of Pathology, Yale University, New Haven, CT, USA. [10]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA. [11]Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. [12]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA. [13]Department of Neuroscience, Yale University, New Haven, CT, USA.
***Corresponding author. Email: drweinberger@libd.org (D.R.W.); flora.vaccarino@yale.edu (F.M.V.); abyzov.alexej@mayo.edu (A.A.)**
†These authors contributed equally to this work.
‡Present address: Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA.
§The Brain Somatic Mosaicism Network authors and affiliations are listed in the supplementary materials.

Results

Mutation discovery and counts across cohorts

For each frozen postmortem brain, one to two regions (cortex, striatum, or hippocampus) were extracted (~1 cm^3) at one of three institutions: Yale University, the Lieber Institute for Brain Development (LIBD), or Harvard University [the data were reported previously (*6*)]. DNA extracted from bulk samples, from fluorescence-activated cell sorted neuronal and glial cell fractions, or from both was analyzed by whole genome sequencing (WGS) on an Illumina platform at various depths, resulting in a combined coverage of >200× per brain for 92% of the samples, and coverage as high as 620× (fig. S1 and table S1). The data were uniformly processed—reads were aligned and somatic point mutations were called in all samples using a previously verified workflow (*7*). Somatic mutations were distinguished from inherited variations on the basis of both their frequency and the lack of overlap with the catalogs of known germline variations in the human population.

In brains with bulk samples sequenced to at least 200× coverage (tier 1 set), we discovered 20 to 60 somatic single-nucleotide mutations per brain with typical allele frequencies ranging from 1 to 10% across all cohorts (Fig. 1A, fig. S2B, and table S2). Additional mutations (included in the tier 2 set) were discovered when considering additional sequencing coverage including data for cell fractions (see methods in the supplementary materials), but for an unbiased comparison across cohorts, we used the tier 1 set unless otherwise noted. In every cohort, there was at least one brain with an outlier high count (i.e., above the upper outer fence of 101) of somatic mutations, which we refer to as a hypermutable brain (Fig. 1A and methods). The mutation calls in those brains had the same substitution spectrum as for other brains (Fig. 1, B and C, and fig. S2E). From read-backed phasing of calls to personal DNA haplotypes, the proportions of supported (i.e., assigned to a single haplotype) and unsupported (i.e., with evidence of being at two haplotypes) calls in the hypermutable brains were the same as in other brains (Fig. 1D), indicating the same quality of calls across all brains. Thus, we concluded that calls in hypermutable brains represent bona fide somatic mutations. Furthermore, we conducted mutation validation in two hypermutable brains, NC7 and LIBD82. In brain NC7, at least 87.5% of mutation calls were validated using single-cell sequencing (Fig. 2). Similarly, we validated at least 80% of mutations in transcriptome in brain LIBD82 using RNA sequencing (methods).

When we excluded the hypermutable brains, there were no significant differences in somatic mutation burden between diseased and normal brains, the allele frequency of the somatic mutations, or the somatic mutation spectrum (Fig. 1, C and E). The combined mutation spectrum was dominated by C>T transitions in CpG motifs and matched the spectrum of mutations arising during development (Fig. 1, B and C, and fig. S2, E and G) (*1*, *7*, *8*). Brains in each cohort had different age distributions, but the mutation burden in nonhypermutable brains did not correlate with age (fig. S2F), which is also consistent with somatic mutations being of early developmental origin. In contrast, the proportion of hypermutable brains rose with age ($P = 8.2 \times 10^{-3}$ by χ^2 test for trend), suggesting that those brains have

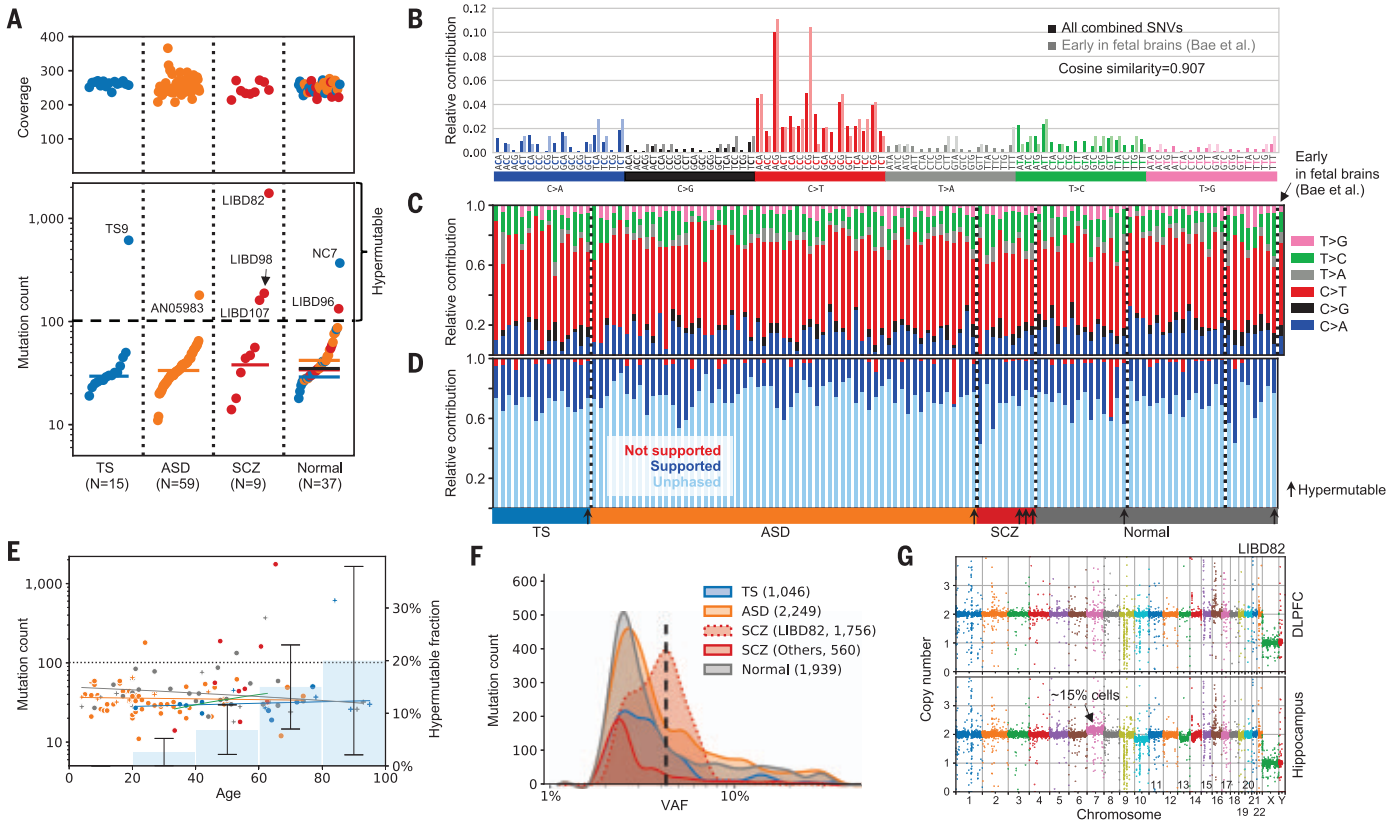


Fig. 1. Mutations discovered across cohorts of brains. (**A**) Summary of coverage and detected somatic point mutations across cohorts using bulk brain samples. Mutation burden across cohorts was comparable when excluding hypermutable brains with >101 mutations (named in the plots). Different colors indicate the institutions that processed the brain samples: Yale University, blue; Harvard University, orange; and LIBD, red. Colored bars show the median mutation count per disease cohort. The black bar represents the median mutation count in all normal brains. (**B**) The combined mutation spectrum is dominated by C>T transitions in CpG motifs and matches the spectrum of developmental mutations (fig. S2). (**C**) The substitution spectra of detected mutations across all brains are comparable. (**D**) Fractions of supported and unsupported calls by assigning to haplotypes are comparable across brains, indicating the same accuracy of calls for hypermutable (indicated by arrows) and nonhypermutable brains. Owing to the short DNA fragments (~450 base pairs), only ~20% of calls can be assigned to a haplotype using nearby heterozygous SNPs; other calls are indicated as unphased. (**E**) When excluding hypermutable samples, the mutation burden does not correlate with age. However, there are more hypermutable brains (those are above the dotted line) in older brains (blue histogram with error bars in the background). Each data point is an individual brain with colors representing phenotypes: TS, blue; ASD, orange; SCZ, red; and normal, gray. Circles indicate males, crosses indicate females. (**F**) Distributions of mutation counts and allele frequencies across cohorts. The numbers in parentheses are the mutation counts. Brain LIBD82 is an exception; its mutations had higher frequencies than those in other brains. The VAF value of the vertical dashed line corresponds to ~15% of aneuploid cells in the hippocampus of that brain (which corresponds to displayed average VAF of ~4% between the cortex and hippocampus). (**G**) Brain LIBD82 had apparent aneuploidies in ~15% of cells in the hippocampus, which may be a signature of incipient glioma or glioblastoma. DLPFC, dorsolateral prefrontal cortex.

mutations arising both in development and during aging (Fig. 1E). Specifically, hypermutable brains constituted 16% of brains over 60 years old, but only 2% of those under 40 years old. One of the hypermutable SCZ brains with the highest mutation burden, LIBD82, had genomic aneuploidies in the hippocampus, including a duplication of chromosome 7 and a deletion of chromosome 10 (Fig. 1, F and G), which is a signature of glioblastoma (*9*). Given that clinical characterization of the brain did not provide any evidence of the disease, the presence of aneuploidies in this brain is consistent with the idea that driver mutations often precede cancer diagnosis by years, if not decades (*10*). Thus, hypermutability in this brain may reflect incipient glioblastoma.

Origin of hypermutability

Hypermutable brains were found in patients and controls and in every cohort of samples generated by each lab. Thus, it is unlikely that the large mutation load in these brains is associated with disease status or arises from cohort- or lab-specific conditions used for brain storage, handling, or sample processing. We hypothesize that the hypermutability arises from individual-specific biological causes. One possibility is that cells of such brains are intrinsically hypermutable. However, the frequency spectrum of mutations in hypermutable brains as compared with nonhypermutable brains was depleted for higher frequency mutations [>6% variant allele frequency (VAF)], overrepresented for intermediate frequencies (3 to 6%), and almost the same for low frequency mutations (<3% VAF) (fig. S2, B to D). These data suggest that intrinsic hypermutability, if it exists, is not innate (that is, present in all cell divisions throughout development) but rather occurs sometime during the lifetime of that individual.

Out of six hypermutable brains, two, TS9 and NC7, had the same known missense mutation (chr1:115,258,747 C>T) in the *NRAS* oncogene. This mutation has been previously identified in multiple cancers and is deemed to be pathogenic in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (ID: COSV54736383). Furthermore, three hypermutable brains (NC7, TS9, and AN05983) carried 11 mutations predicted to have a damaging effect in genes previously implicated in cancers (Table 1), such as *MTOR*, *TET2*, *DNMT3A*, and *IDH2*. Overall, hypermutable

Table 1. List of detected somatic mutations that putatively affect genes previously implicated in cancers and genes previously associated with neuropsychiatric diseases. Some mutations are also annotated as pathogenetic in the COSMIC database (*33*). A known mutation in *NRAS* oncogene is present in two brains. Validation: A, by amplicon; R, previously reported and validated by Rodin *et al*. (*6*); and S, by single cells. Gene annotation: B, cancer implicated genes in Bailey *et al*. (*11*); C, cancer gene census (*12*); ASD, SFARI genes; and SCZ, compiled from previous SCZ studies (*34*–*37*).

Mutation	Brain	Phenotype	Hypermutable	VAF	Validation	Consequence (VEP)	Gene	Gene annotation	Genomic mutation ID (COSMIC)	Prediction (COSMIC)
In genes implicated in cancer										
chr1:115,258,747 C>T	NC7	Normal	Yes	1.84%	A, S	Missense	*NRAS*	B, C	COSV54736383	Pathogenic
	TS9	TS	Yes	1.10%	A					
chr2:25,466,766 C>G	NC7	Normal	Yes	1.91%	A, S	Splice	*DNMT3A*	B, C	–	–
chr3:81,695,552 G>T	TS9	TS	Yes	1.26%	n/a	Missense	*GBE1*	–	COSV101450116	Pathogenic
chr4:106,156,747 C>T*	TS9	TS	Yes	1.05%	A	Stop	*TET2*	B, C	COSV54395664	Neutral
chr4:106,197,392 G>T			Yes	1.75%	A	Stop			–	–
chr4:183,267,873 C>T	TS9	TS	Yes	1.27%	n/a	Missense	*TENM3*	–	COSV69301741	Pathogenic
chr12:48,107,157 C>G	TS9	TS	Yes	1.16%	n/a	Missense	*ENDOU*	–	COSV57465821	Pathogenic
chr15:90,631,934 C>T*	NC7	Normal	Yes	3.55%	S	Missense	*IDH2*	B, C	COSV57468751	Pathogenic
chrX:129,168,460 C>T	NC7	Normal	Yes	2.99%	S	Stop	*BCORL1*	C	COSV54390206	Neutral
chr1:11,300,357 T>C	AN05983	ASD	Yes	2.63%	A	Splice	*MTOR*	B, C	–	–
chr10:32,326,503 A>T*	LIBD99	Normal	No	2.62%	n/a	Missense	*KIF5B*	C	–	–
chr6:168,315,889 C>T*	TS1	TS	No	0.95%	A	Stop	*MLLT4*	C	COSV60045383	Pathogenic
chr4:1,977,081 G>A	UMB797	ASD	No	2.78%	n/a	Missense	*WHSC1*	B, C	COSV56389769	Pathogenic
chr7:148,495,674 A>G	UMB5297	ASD	No	2.99%	n/a	Missense	*CUL1*	B	–	–
In disease-implicated genes in brains with diseases										
chr1:11,300,357 T>C	AN05983	ASD	Yes	2.63%	A	Splice	*MTOR*	ASD	–	–
chr2:179,612,387 T>A	UMB1638	ASD	No	2.09%	n/a	Stop	*TTN*	ASD	–	–
chr8:146,033,199 C>T	AN13654	ASD	No	4.15%	n/a	Missense	*ZNF517*	ASD	–	–
chr10:55,600,099 C>T	UMB4231	ASD	No	4.41%	A	Missense	*PCDH15*	ASD	COSV57387920	Pathogenic
chr16:20,976,490 C>T	AN09412	ASD	No	12.09%	R	Missense	*DNAH3*	ASD	–	–
chr2:73,800,041 T>G*	LIBD100	SCZ	No	4.46%	n/a	Missense	*ALMS1*	SCZ	–	–

*Tier 2 mutation (see methods).

brains were enriched for damaging mutations in genes implicated in cancers (*11*, *12*) as compared with nonhypermutable brains ($P = 2.4 \times 10^{-3}$ by Fischer's exact test; fig. S3A).

Given the above data, we hypothesized that the cell lineage carrying one or more of these mutations underwent a clonal expansion, which allowed other preceding lineage-specific mutations to rise to a detectable frequency. In both brains TS9 and NC7, virtually all of the mutations were present in both bulk cortex and striatum at an allele frequency of ~2%, as well as in multiple other cell fractions at various frequencies (Fig. 2, A and B, and fig. S3, B and C). This could be explained by clonal hematopoiesis coupled with blood cell infiltration into the brain, such that mutations from blood are present in different brain regions. Hence, we investigated the expression of blood-specific genes (i.e., markers) in the studied brains. Several markers consistently indicated that only brain NC7 had a higher fraction of blood cells as compared with all other brains (fig. S4). However, in NC7, all the mutations were present in sorted interneurons from the striatum (STR-INT) at a much higher frequency (15 to 25% VAF), suggesting that the expanded lineage originated not in the blood but rather in the embryonic basal ganglia, where interneurons are generated, and then populated the cortex and striatum by cell migration.

We tested the hypothesis of lineage expansion (either of interneurons or of hematopoietic cells) by sequencing 16 single nuclei isolated from a different sample of the STR-INT fraction from brain NC7. Of all the somatic mutations discovered in bulk NC7 samples, 346 (94%) were genotyped in 8 of 16 nuclei, with 154 to 197 mutations genotyped in each nucleus (Fig. 2C). The genotyping rate, 42 to 53% across nuclei, was comparable to the genotyping sensitivity (34 to 46%; see methods) allowing us to extrapolate that each of the eight nuclei contained almost all the mutations discovered in the bulk cortex and striatum and proving that the mutations belonged to the same cell lineage.

Unlike in brains NC7 and TS9, mutations in other hypermutable brains with data available for multiple regions (LIBD82, LIBD96, LIBD98, and LIBD107) were restricted to one brain region (fig. S3, D to F). For brain LIBD82, the frequencies (~15% in the hippocampus) of point mutations and of genomic aneuploidies were matching, consistent with both mutation types being present in expanded lineages (Fig. 1, E and F). Altogether, the above evidence suggests that hypermutability in some brains is caused by lineage expansion, either during development or later in life.

Somatic mutations reveal uneven cell lineage distribution in the brain

For each brain processed at Yale University (19 TS and 21 neurotypical brains), we obtained sequencing data for bulk cortex, striatum, and up to eight cell fractions per brain region (see methods). In both the cortex and striatum, we successfully fractionated neurons and nonneuronal cells ($NeuN^+/NeuN^-$); in the striatum, we also fractionated medium spiny neurons (STR-MSN) and interneurons (STR-INT) (fig. S1B and methods). From each brain processed at LIBD, we obtained sequenced data for bulk cortex and hippocampus. The frequency of mosaic mutations in multiple regions and

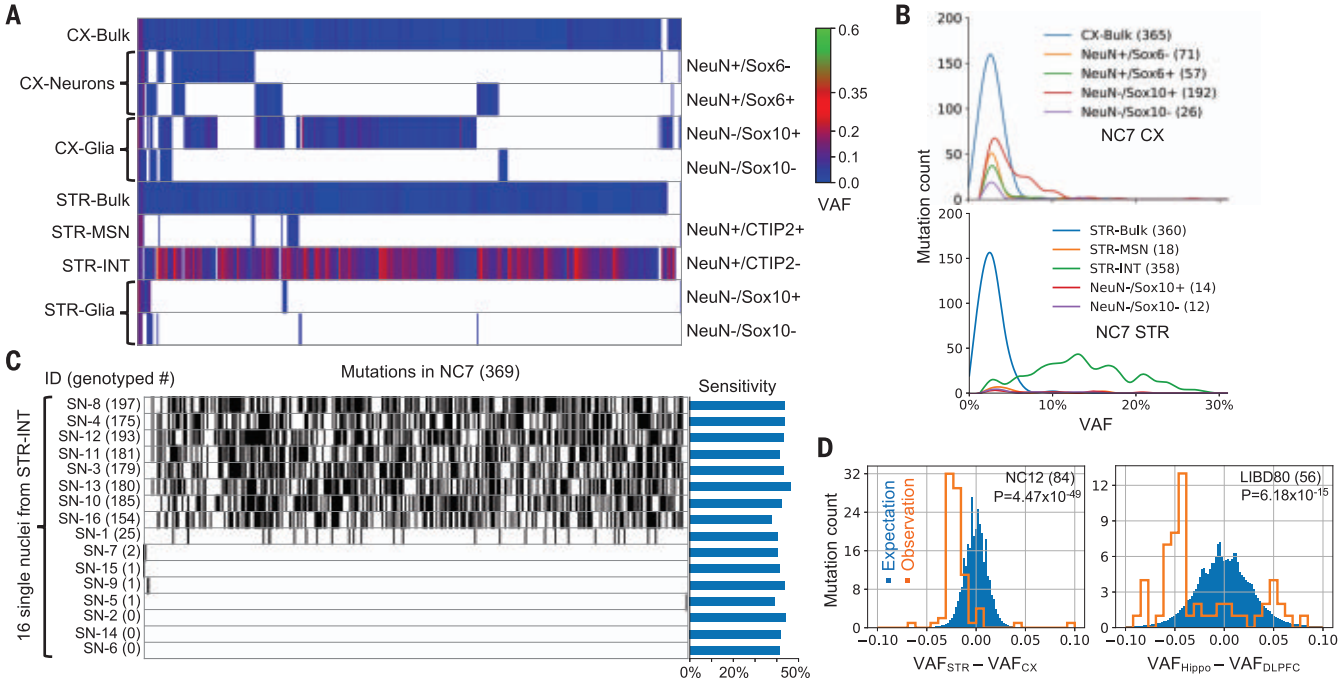


Fig. 2. Uneven cell lineage distribution of somatic mutations in brain. (**A**) Mutation allele frequencies in samples from brain NC7. Almost all mutations discovered in NC7 are present in the striatal interneuron fraction (STR-INT) at high VAFs. CX, cortex. (**B**) Distribution of mutation allele frequencies across samples in NC7. (**C**) Genotyping of mutations in 16 single nuclei originated from the NC7 STR-INT fraction by sequencing at 5× depth. Black bars represent genotyped mutations. (**D**) Examples of brains where frequencies of mutations are significantly biased toward the cortex as compared with the striatum (NC12) or as compared with the hippocampus (LIBD80).

fractions allowed us to explore the cell lineage distribution in the brain.

To investigate whether the above result (Fig. 2, A and B) could reflect a more general phenomenon, we tested for the difference in mutation VAFs between the cortex and striatum. If the early lineage distribution is the same across these two regions, there should be, on average, no difference in VAFs between them (methods). Out of 22 brains of the Yale cohort for which we had a complete dataset for the bulks and fractions, eight (six nonhypermutable) showed a statistically significant difference between the cortex and striatum (Fig. 2D and fig. S5, A and B). Except for the hypermutable brain NC7, a higher VAF of somatic mutations was always observed in the cortex. Even in NC7, the VAFs were higher in the cortex when excluding neuronal fractions (Fig. 2A and fig. S5C). Similarly, in 5 out of 13 nonhypermutable brains processed by LIBD (six SCZ and seven neurotypical brains), we also observed a higher VAF in the cortex as compared with the hippocampus (Fig. 2D and fig. S5, D and E). Combining the brains with biased VAFs from the two datasets, we observed a significant overrepresentation of brains with higher VAF values in the cortex ($P = 5 \times 10^{-4}$ by binomial test). A possible explanation of this observation is that the founder population of the cortex is allocated with fewer earlier lineages, and so the frequency of each lineage is, on average, higher. Alternatively, the cortex may have a higher propensity for expansion in some lineages when compared with the striatum and hippocampus.

Relevance of somatic mutations to genome function

Given the mostly random distribution of somatic mutations in the human genome, the majority (~60%) of the mutations were outside genes with roughly the same distribution across the genome in each cohort (Fig. 3A and fig. S6A). Approximately 2 to 3% of mutations were in the coding part of the genome, and another 6% had the potential to affect regulatory regions. Six putative deleterious mutations in genes previously associated with neuropsychiatric diseases were found in ASD and SCZ brains (Table 1). For example, in the ASD brain UMB4231, we detected and validated a missense mutation in *PCDH15*—a gene encoding a protein mediating calcium-dependent cell–cell adhesion, which was previously associated with ASD (*13*). Furthermore, in the ASD brain AN05983, we detected and validated a splice mutation in *MTOR*, a gene that is part of the insulin-like growth factor 1–phosphatidylinositol 3-kinase (IGF1-PI3K) signaling pathway implicated in ASD (*14*).

We did not detect deleterious somatic mutations in TS-associated genes, but a missense mutation in the *ARHGEF6* gene (chrX:135,761,734 G>T) was found in brain TS9 at a relatively high VAF in both the cortex (4.2% VAF) and striatum (1.6% VAF). The *ARHGEF6* gene has been associated with X-linked intellectual disability, dendrite orientation, and cell polarity (*15*). Indeed, knockout of *ARHGEF6* in mouse models results in hippocampal dendrite and/or synaptic abnormalities and deficits in learning (*16*). Deficits in cell polarity have been implicated in Tourette syndrome from analyses of rare de novo mutations (*17*).

We next investigated the possible functional impact of the discovered mutations in nonhypermutable brains on enhancer-like elements active in the fetal human brain (*18*) and in an organoid model of the developing human brain (methods). For each set of enhancer-like elements, analysis of the ASD cohort revealed an excess, as compared with controls, of somatic mutations creating putative transcription factor (TF) binding motifs ($P < 10^{-4}$ by binomial test) (Fig. 3B and fig. S6C). Top-ranked mutations in ASD brains created 16 putative binding motifs for the myeloid ectopic viral integration site 1 (MEIS1), MEIS2, and MEIS3 transcription factors (Fig. 3C); in normal brains, only four such mutations were detected (not shown in Fig. 3C). Those TFs were also among the top-ranked when simulating random distribution of mutations in ASD brains across the genome (not shown in Fig. 3C).

Somatic structural mutations

To call structural mutations, we used an approach implemented in CNVpytor (Fig. 4). This

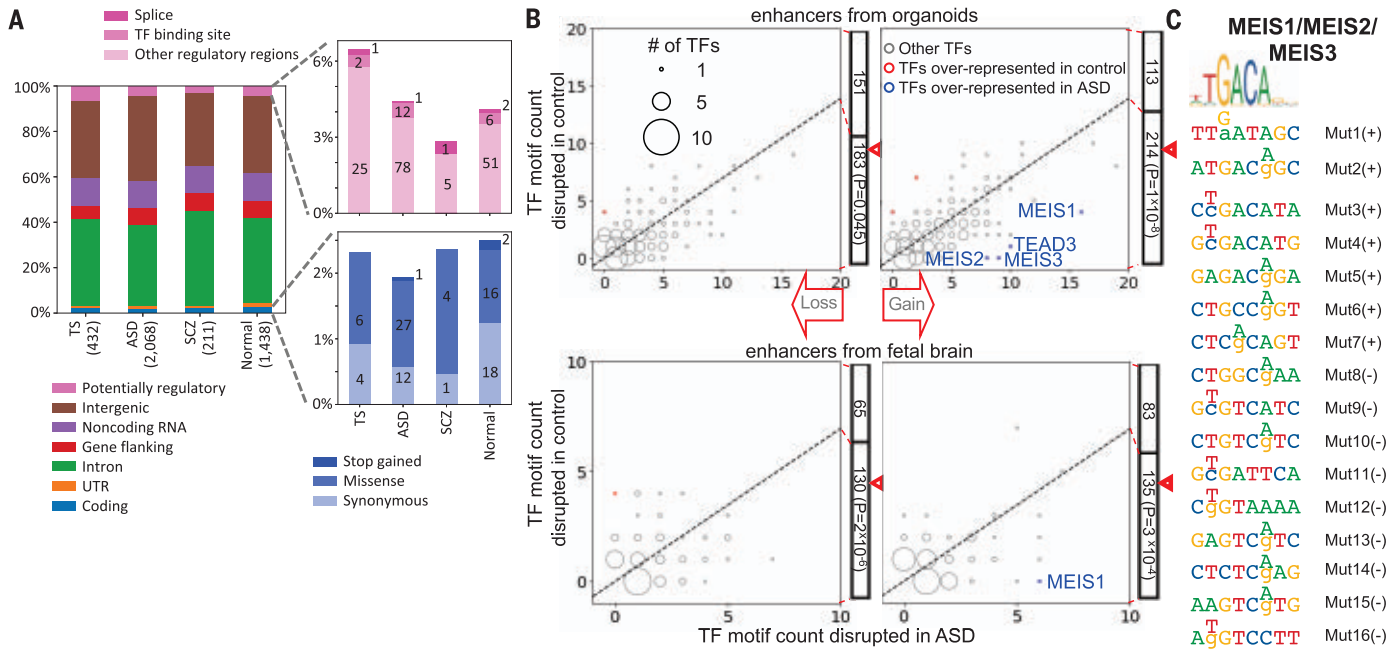


Fig. 3. Relevance of somatic mutations to genome function. (**A**) In each cohort, mutations from nonhypermutable brains have roughly the same distribution across genomes and similar predicted functional impacts by Variant Effect Predictor (VEP). UTR, untranslated region. (**B**) Counts of putative TF motifs disrupted by somatic mutations within enhancer-like elements from organoids (upper plots) or fetal brain (lower plots), showing that mutations significantly disrupt more putative TF binding sites in ASD versus controls. Each circle represents the number of TFs with *x* (count of mutations in ASD) and *y* coordinates (count of mutations in controls) of the circle. (**C**) Somatic mutations in the ASD cohort predicted to lead to gain of binding sites for MEIS1, MEIS2, and MEIS3 in enhancer-like elements. The consensus motif is on top. Mutations on the positive (+) or negative (−) strand are enumerated on the right, with coordinates listed in table S3. For each mutation, a reference base is shown with a lowercase letter and the mutation base is noted by a capital letter above it.

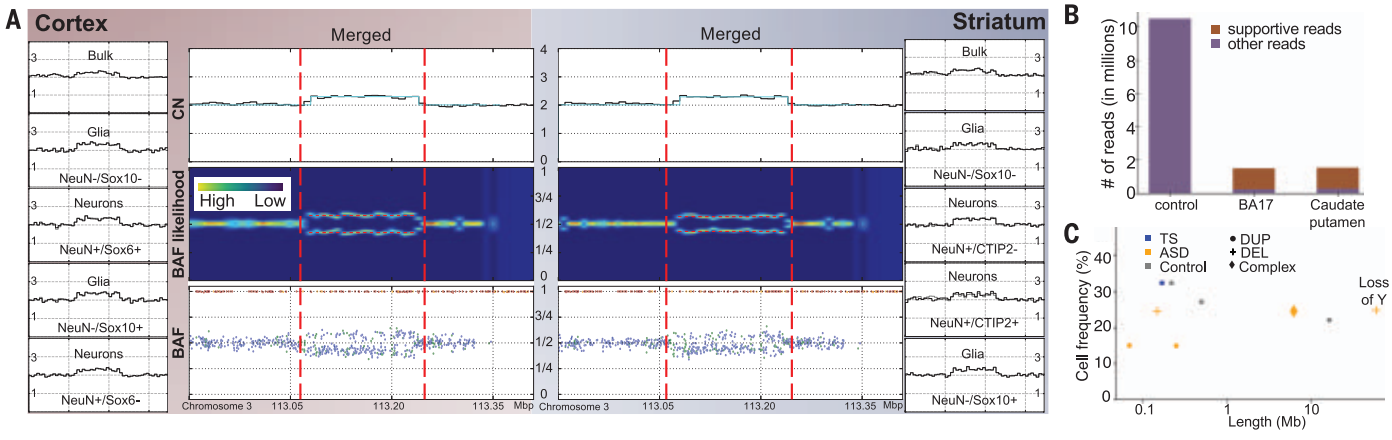


Fig. 4. Detection and validation of large somatic structural mutations. (**A**) Example of a duplication on chromosome 3 (chr3:113,067,261–113,233,476) identified in brain TS1 in eight cell fractions and two bulks, the cortex and striatum. Red dashed vertical lines outline the region of the duplication. The two merged plots combine data for bulk and fractions for the cortex (left) and striatum (right) and present an increase in the read depth corresponding to the cell frequency of ~32% (top), split in B allele frequency (BAF) likelihood function (brighter color corresponds to higher values) matching to the same cell frequency (middle), and BAF for individual SNPs (bottom). SNPs in the accessible genome (P-bases) are in blue, other SNPs are in green. (**B**) Amplicon sequencing validation of the duplication. Bar plot shows the number of reads mapping to the junction of the duplication in the cortical region BA17 and caudate putamen of TS1 and in control sample. DUP, duplication; DEL, deletion; LOH, loss of heterozygosity. (**C**) Summary of size and cell frequency of all structural mutations detected from all analyzed individuals.

tool segments the genome, considering both depth of coverage and split in allele frequencies for single-nucleotide polymorphisms (SNPs), allowing us to call deletions, duplications, and copy number neutral losses of heterozygosity (CNN-LOH). The called mutations can be either inherited or somatic, and we distinguished between these two types by their cell frequency (methods). We excluded brain LIBD82 for having aneuploidies and a few other brains because of poor coverage and data quality. We focused on a total of nine high-confidence calls encompassing an entire chromosome or for which we could resolve breakpoints (tables S4 and S5 and figs. S7 to S21). For five mutations,

we attempted validation in brain tissue by polymerase chain reaction across the breakpoints, which confirmed all of them (Fig. 4B and fig. S7). Two more mutations were previously validated (*19*).

In total, we detected nine rearrangements in nine nonhypermutable brains (7% of all brains) ranging from 70 kilo–base pairs to 16.5 mega–base pairs in length, with the largest mutation representing the loss of chromosome Y. That chromosome loss and another deletion were detected only in ASD brains. Seven somatic structural mutations were duplications, which were found in ~5% of both normal and diseased brains. The two duplications discovered in the Tourette cohort (brains TS1 and NC6, the only brains for which more than one region was available) were present in all bulk samples and fractions analyzed (Fig. 4A and fig. S7A). This finding is consistent with the idea that all these duplications occurred during development, which is also supported by their frequency in at least 10% of cells. Consistent with a developmental origin, sequence microhomologies (1 to 4 base pairs) were found in most of the resolved breakpoints (fig. S22), suggesting that the corresponding mutations were generated by replicative mechanisms, as previously observed for mosaic copy number alterations in the fetal human brain (*20*). In many aspects, duplications resemble de novo copy number variants (CNVs) in the CNV mutator phenotype (*21*)—in that they are mostly tandem duplications of early developmental origin and have sequence microhomologies at breakpoints—although the mutations detected in this study were smaller than those reported in previous studies. Thus, we hypothesize that somatic duplications represent background mutations that arise in early development, with the CNV mutator phenotype representing the extreme manifestation of this phenomenon leading to a disease phenotype.

Discussion

Our study reveals a complex etiology of somatic mutations in the brain and their possible relationship with phenotype. While most of the studied brains had just a few dozen mutations of likely developmental origin in proliferating cells, ~6% of brains were classified as hypermutable. Hypermutability was not related to disease but did increase with age, reaching at least a 3% population frequency (95% confidence interval) for brains over 40 years old. The exact cause of hypermutability is yet to be determined, but besides a possible higher mutation rate, in some brains it originates from expansion of a cell lineage. Given the association of hypermutability with age and its frequent localization in one brain region, we hypothesize that it could be related to local clonal gliogenesis in adult brains, consistent with previous reports of increased glial fraction with age (*22*, *23*). Such gliogenesis could be relevant to cognitive aging, such as Parkinson's and Alzheimer's disease. Alternatively, hypermutability could be a remnant of a brain tumor escaped earlier in life.

Half of the hypermutable brains studied in this work carried mutations in genes that are involved in clonal hematopoiesis in aging adults (*24*), consistent with previous observations from the analysis of a panel of genes (*25*). Thus, it is tempting to speculate that increases in clonal hematopoiesis with age (*24*), coupled with a leakier blood-brain barrier with age and extravasation of blood in brain tissue (*26*), could explain the higher mutation burden in those brains. Analysis of expression data in brain NC7 is consistent with an increased fraction of blood cells in that brain. However, analysis of copy number for loci of immunoglobulin receptors in B and T cells does not reveal an obvious presence of immune cells in the clonally expanded lineage (fig. S23). Furthermore, in that brain, all mutations were present in two regions and were enriched in postmitotic neuronal cell fractions, lending support to yet another possibility, that of a neural lineage clonal expansion during development. In support of this scenario, a spectrum of de novo variants in *DNMT3A*, which overlaps with those found in hematologic malignancies, causes Tatton-Brown-Rahman syndrome, a developmental disorder with macrocephaly (*27*). It was also recently shown that human interneurons, in contrast to other neurons, continue to expand until birth (*28*). Nonetheless, establishing the cause(s) and extent of the phenomenon of hypermutability requires analysis of larger sets of brains and additional analyses at the single-cell level.

Somatic structural mutations were detected in ~7% of brains, but because most of them were duplications (in ~5% of brains), they are unlikely to have functional consequences. The ASD cohort was the only one where deletions were detected. Additional CNV calls (without resolved breakpoint but with confirmed haplotype imbalance) were also made only in the ASD cohort (table S4), perhaps pointing to the relevance of somatic deletions for the ASD phenotype. That said, further studies of larger ASD and other disease cohorts are necessary to make definitive conclusions on this observation.

In ASD brains, somatic point mutations in noncoding regions create putative TF binding motifs within enhancer-like elements that are active in the developing human brain. We hypothesize that some of the newly created motifs result in binding of their corresponding TFs, thereby affecting TF protein dosage and dysregulating gene regulatory networks. The top-ranked affected motifs are putative binding sites for MEIS genes, which are homeodomain-containing transcriptional activators that promote chromatin decompaction (*29*). MEIS TFs have been implicated in intellectual disabilities; act as cofactors of HOX genes; and are regulators of proliferation, growth, neurogenesis, and patterning during development (*29*). MEIS2 is highly expressed in cortical development and was defined as a marker of a subpopulation of cortical interneurons that populates the white matter (*30*). We previously proposed that deleterious de novo ASD variants are enriched in homeodomain TF binding motifs within developmental enhancer-like elements (*31*). Similarly, enhancer-like elements associated with ASD genes from the Simons Foundation Autism Research Initiative (SFARI) collection are enriched for homeobox binding motifs (*32*). Therefore, we suggest that functional studies of mutations within homeodomain binding motifs during human development might aid efforts in understanding the etiology of ASD.

REFERENCES AND NOTES

1. T. Bae *et al*., *Science* **359**, 550–555 (2018).
2. L. Fasching *et al*., *Science* **371**, 1245–1248 (2021).
3. M. A. Lodato *et al*., *Science* **359**, 555–559 (2018).
4. F. Abascal *et al*., *Nature* **593**, 405–410 (2021).
5. M. J. McConnell *et al*., *Science* **356**, eaal1641 (2017).
6. R. E. Rodin *et al*., *Nat. Neurosci.* **24**, 176–185 (2021).
7. Y. Wang *et al*., *Genome Biol.* **22**, 92 (2021).
8. S. Bizzotto *et al*., *Science* **371**, 1249–1253 (2021).
9. I. Crespo *et al*., *J. Mol. Diagn.* **13**, 634–647 (2011).
10. M. Gerstung *et al*., *Nature* **578**, 122–128 (2020).
11. M. H. Bailey *et al*., *Cell* **173**, 371–385.e18 (2018).
12. Z. Sondka *et al*., *Nat. Rev. Cancer* **18**, 696–705 (2018).
13. R. Anney *et al*., *Hum. Mol. Genet.* **21**, 4781–4792 (2012).
14. W. Zhang *et al*., *Genes Dev.* **34**, 580–597 (2020).
15. M. Meseke, G. Rosenberger, E. Förster, *Eur. J. Neurosci.* **37**, 1404–1412 (2013).
16. G. J. A. Ramakers *et al*., *Hum. Mol. Genet.* **21**, 268–286 (2012).
17. S. Wang *et al*., *Cell Rep.* **24**, 3441–3454.e12 (2018).
18. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, *Nature* **590**, 300–307 (2021).
19. M. A. Sherman *et al*., *Nat. Neurosci.* **24**, 197–203 (2021).
20. S. Sekar *et al*., *Genome Res.* **30**, 1695–1704 (2020).
21. P. Liu *et al*., *Cell* **168**, 830–842.e7 (2017).
22. J. R. Conde, W. J. Streit, *J. Neuropathol. Exp. Neurol.* **65**, 199–203 (2006).
23. T. G. Beach, R. Walker, E. G. McGeer, *Glia* **2**, 420–436 (1989).
24. F. Zink *et al*., *Blood* **130**, 742–752 (2017).
25. M. J. Keogh *et al*., *Nat. Commun.* **9**, 4257 (2018).
26. W. A. Banks, M. J. Reed, A. F. Logsdon, E. M. Rhea, M. A. Erickson, *Nat. Aging* **1**, 243–254 (2021).
27. W. Shen *et al*., *Am. J. Med. Genet. A.* **173**, 3022–3028 (2017).
28. M. F. Paredes *et al*., *Science* **375**, eabk2346 (2022).
29. D. Schulte, D. Geerts, *Development* **146**, dev174706 (2019).
30. S. Frazer *et al*., *Nat. Commun.* **8**, 14219 (2017).
31. A. Amiri *et al*., *Science* **362**, eaat6720 (2018).
32. A. E. Trevino *et al*., *Science* **367**, eaay1645 (2020).
33. J. G. Tate *et al*., *Nucleic Acids Res.* **47**, D941–D947 (2019).
34. E. Rees *et al*., *Nat. Neurosci.* **23**, 179–184 (2020).
35. D. P. Howrigan *et al*., *Nat. Neurosci.* **23**, 185–193 (2020).
36. A. F. Pardiñas *et al*., *Nat. Genet.* **50**, 381–389 (2018).
37. Schizophrenia Working Group of the Psychiatric Genomics Consortium, *Nature* **511**, 421–427 (2014).
38. F. M. Vaccarino *et al*., BSMN–Commonly Processed Data #814, NIMH Data Archive (2022); https://doi.org/10.15154/1506068.
39. Y. Wang *et al*., bsmn/bsmn-pipeline: Comprehensive identification of somatic nucleotide variants in human brain tissue, version 2, Zenodo (2020); https://doi.org/10.5281/zenodo.4321679.
40. T. Bae *et al*., Analysis of somatic mutations in 131 human brains reveals aging associated hypermutability, version 1, Zenodo (2022); https://doi.org/10.5281/zenodo.6374179.

ACKNOWLEDGMENTS

We thank the Tourette's Association of America for collecting and providing brain tissue. We also thank the donors and their families. We acknowledge the NIH NeuroBioBank and its director, S. Berretta, for providing postmortem tissue from TS and NC individuals. We are grateful to G. Lyon of the Yale Flow Cytometry Facility for help with fluorescence-activated cell sorting of brain postmortem nuclei. Some tissues were obtained from Autism

BrainNet, a resource of the SFARI. Autism BrainNet also manages the Autism Tissue Program (ATP) collection, previously funded by Autism Speaks. We are grateful and indebted to the families who donated tissue for research purposes to Autism BrainNet and the ATP. We thank members of the Brain Somatic Mosaicism Network (BSMN) consortium for useful discussions and suggestions during the study and preparation of the manuscript. Members of the BSMN consortium are listed in the supplementary materials. We express our gratitude to the National Institute of Mental Health Data Archive and Synapse teams for providing collaborative data storage space and support for data transfer needed for the successful completion of this study. **Funding:** This work was supported by National Institutes of Health grants U24CA220242, U01MH106892, U01MH106876, U01MH106883, U01MH106874, U01MH106893, U01MH106882, U01MH108898, U01MH106891, U01MH106884, and R01MH109648 and by Simons Foundation Autism Research Initiative grant 399558. The Urban laboratory receives funding through the Jaswa Innovator Award and from B. Blackie and W. McIvor. **Ethics statement:** The analyzed samples and generated data were derived from deidentified postmortem individuals. Tourette syndrome brains sequenced by Yale University were obtained from the Harvard Brain and Tissue Resource Center (HBTRC). Most normal brains sequenced at Yale University were obtained from NIH NeuroBioBank. For normal brains from Yale University, autopsy authorization, consent to utilize tissues for research, and consent to publish were provided by the patients' next of kin according to Connecticut state law and approved protocols of the Yale University BioBank. The research study was approved by the Yale Alzheimer's Disease Research Center (ADRC) and was reviewed and deemed exempt by the Yale University Institutional Review Board. The BioBank protocols are in accordance with the ethical standards of Yale University. Handling of brains originating from and sequenced by the Lieber Institute was conducted under the approved WCG IRB protocol 20111080 titled "Collection of Postmortem Human Brain, Blood and Scalp Samples for Neuropsychiatric Research." Data for brains sequenced by Harvard University are publicly available from NIMH Data Archive. **Author contributions:** J.F.L. performed retrospective clinical phenotyping of the TS cases by medical record and interviews with family members. L.F., R.C.R., A.H., J.E.K., and T.M.H. processed brains. L.F., J.H.S., J.M., and A.J. performed experiments and generated the data. T.B., Y.W., J.H.S., M.S., Y.J., S.N., F.W., and A.P. performed data analyses. L.F., C.D., R.P., Y.C., and R.Y. conducted validation of mutations. T.B., L.F., Y.W., and M.S. prepared display items. T.B., L.F., Y.W., M.S., S.N., C.D., R.E.S., D.R.W., F.M.V., and A.A. wrote the initial draft of the manuscript. T.M.H., R.E.S., C.A.W., A.E.U., D.R.W., F.M.V., and A.A. oversaw experiments and analyses. All authors participated in discussions of results and manuscript editing. **Competing interests:** C.A.W. is a paid consultant (cash, no equity) to Third Rock Ventures and Flagship Pioneering (cash, no equity) and is on the clinical advisory board (cash and equity) of Maze Therapeutics. No research support is received. These companies did not fund and had no role in the conception or performance of this research project. D.R.W. is on the scientific advisory boards of Sage Therapeutics and Pasithea Therapeutics. **Data and materials availability:** WGS data and call sets have been deposited in the NIMH Data Archive (NDA Study ID 814) (*38*). Data for brain organoids are available at NIMH Data Archive (NDA Collection ID 2424; https://nda.nih.gov/edit_collection.html?id=2424). The BSMN mutation calling workflow is available on GitHub (https://github.com/bsmn/bsmn-pipeline) and Zenodo (*39*). CNVpytor is available on GitHub (https://github.com/abyzovlab/CNVpytor) and on bio.tools (ID: cnvpytor). The code for data analysis is freely available on Zenodo (*40*). All other data are in the main paper or supplementary materials.
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6222
Materials and Methods
Figs. S1 to S26
Tables S1 to S7
The Brain Somatic Mosaicism Network Authors and Affiliations
References (*41*–*60*)
MDAR Reproducibility Checklist

Submitted 4 October 2021; accepted 20 June 2022
10.1126/science.abm6222

REPORTS

BIOIMAGING

Bioadhesive ultrasound for long-term continuous imaging of diverse organs

Chonghe Wang[1]†, Xiaoyu Chen[1]†, Liu Wang[1], Mitsutoshi Makihata[2], Hsiao-Chuan Liu[3], Tao Zhou[1], Xuanhe Zhao[1,4]*

Continuous imaging of internal organs over days could provide crucial information about health and diseases and enable insights into developmental biology. We report a bioadhesive ultrasound (BAUS) device that consists of a thin and rigid ultrasound probe robustly adhered to the skin via a couplant made of a soft, tough, antidehydrating, and bioadhesive hydrogel-elastomer hybrid. The BAUS device provides 48 hours of continuous imaging of diverse internal organs, including blood vessels, muscle, heart, gastrointestinal tract, diaphragm, and lung. The BAUS device could enable diagnostic and monitoring tools for various diseases.

Wearable devices that can continuously monitor human physiology in a noninvasive manner represent a pivotal trend in precision and digital medicine (*1*). Particularly, deep-tissue physiology, including internal organ signals and dynamics, contains crucial information regarding health and diseases (*2*). Although current wearable devices have successfully recorded physical and chemical signals from the skin such as electrocardiogram data (*3*–*5*) and sweat metabolites and electrolytes (*6*–*8*), clinical-grade imaging of various internal organs remains a central task and a challenge in the field of wearable devices. Ultrasound imaging with zero radiation allows clinicians to evaluate tissue and organ functions and diagnose various diseases (*9*). If sampled frequently over days to months, ultrasound imaging of internal organs could help clinicians monitor health status, observe disease progression, and assess disease risks (*10*–*12*); it could also lead to discoveries in and a better understanding of developmental biology (*13*).

Wearable ultrasound devices have the potential for continuous imaging of internal organs. Conventional wearable ultrasound imaging usually relies on mounting bulky ultrasound probes on the skin by either robotic systems (*14*, *15*) or mechanical fixtures such as straps and tapes (*16*–*18*), which hamper patients' mobility and wearing convenience and comfort. In addition, conventional wearable ultrasound imaging is only applicable to limited body parts such as muscles because of the thick ultrasound probe and the high pressure exerted by the probe (*16*). For example, mechanically mounting a conventional ultrasound probe on the neck to image the carotid artery, jugular vein, and vagus nerve could cause suffocation.

Although the stretchable ultrasound imaging device has improved wearability (*12*), it still suffers from limitations, including low imaging resolution, unstable imaging quality during body motions (fig. S1), a short continuous imaging duration (1 hour), and susceptibility to device failure (table S1). (The continuous imaging duration is defined as the longest time that a device can be adhered or worn on the skin.) These limitations mainly stem from the design of the device, which consists of piezoelectric elements over a stretchable substrate (Fig. 1A). Although the stretchable substrate can deform conformally with the skin, it limits the density of elements (156 per square centimeter) and is incompatible with backing and matching layers, leading to a low imaging resolution (*12*). In addition, when the substrate deforms with the skin under body motions, the elements' spatial and angular positions vary unpredictably, which hampers the imaging stability of the device (fig. S1). Furthermore, existing wearable ultrasound devices, in both rigid and stretchable form factors, mostly rely on hydrogel or elastomer couplants for acoustic transmission to the skin (*19*–*22*). However, the hydrogel couplants usually get dehydrated or detached from the skin over a few hours (*19*, *20*), and the elastomer couplants are too damping to image deep organs (*21*, *22*) (fig. S2 and table S2).

We report a bioadhesive ultrasound (BAUS) device that consists of a thin and rigid ultrasound probe robustly adhered on the skin via a couplant layer made of a soft, tough, antidehydrating, and bioadhesive hydrogel-elastomer hybrid (Fig. 1, B to D). The BAUS probe enables a high density of elements (400 elements per

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Makihata Engineering, San Jose, CA 95130, USA. [3]Department of Radiology, Mayo Clinic, Rochester, MN 55905, USA. [4]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
***Corresponding author. Email: zhaox@mit.edu**
†These authors contributed equally to this work.

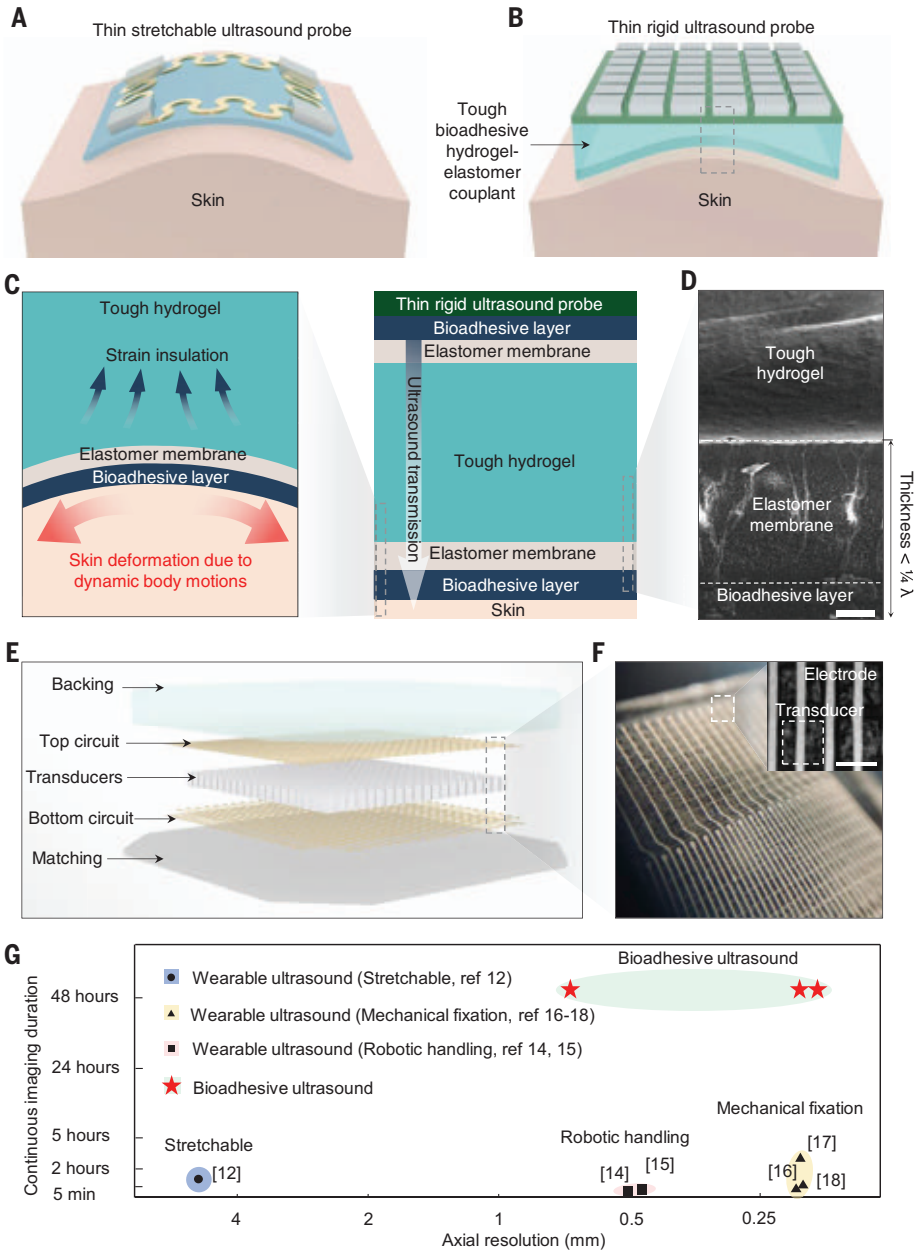


Fig. 1. Design and imaging performances of the BAUS device. (**A**) Existing epidermal wearable ultrasound imaging usually relies on thin and stretchable devices physically attached to the skin. (**B**) The BAUS device consists of a thin and rigid ultrasound probe robustly adhered to the skin via a couplant made of a soft, tough, antidehydrating, and bioadhesive hydrogel-elastomer hybrid. The dashed rectangle indicates the region shown on the right in (C). (**C**) The BAUS couplant consists of a tough hydrogel, containing 90 wt % water, that is encapsulated by an elastomer membrane further coated by a thin bioadhesive layer. The BAUS couplant maintains robust adhesion between the BAUS probe and the skin over 48 hours and insulates the BAUS probe from skin deformation during dynamic body motions. (**D**) Scanning electron microscopy (SEM) image of the cross section of the BAUS couplant. The total thickness of the elastomer membrane and the bioadhesive layer is less than one-fourth of the ultrasound wavelength λ. Scale bar is 10 μm. (**E**) Schematics of the BAUS probe structure. The BAUS probe consists of an array of high-performance piezoelectric elements, which are controlled by the top and bottom circuits. The top and bottom circuits are covered by the acoustic backing and matching layers, respectively. The dashed rectangle indicates the region shown in (F). (**F**) Optical microscopic image of a BAUS probe with 3D-printed top and bottom circuits. The inset image provides a zoomed-in top view of the device showing the top circuit and piezoelectric elements. Scale bar is 0.5 mm. (**G**) Comparison of axial-imaging resolutions and continuous-imaging durations of existing wearable ultrasound imaging devices and the BAUS device.

square centimeter; fig. S3), stable element positions under dynamic body motions, and high reliability of the probe in long-term applications. The BAUS couplant effectively transmits acoustic waves, insulates the BAUS probe from skin deformation, and maintains robust and comfortable adhesion on the skin over 48 hours.

The BAUS probe consists of an array of high-performance piezoelectric elements with a center frequency of 3, 7, or 10 MHz (Fig. 1, E and F; fig. S3; and table S3). Each element is controlled by the top and bottom circuits of the BAUS probe. The circuit can be fabricated with three techniques: three-dimensional (3D) printing (Fig. 1F, fig. S4, and movie S2) (*23*), laser etching (*24*), and photolithography (fig. S5) (*25*), which give circuit-line resolutions of 100, 10, and 1 μm, respectively. The bottom circuit is covered by an optimized acoustic matching layer to enhance the acoustic transmission to the skin, and the top circuit is covered by an optimized acoustic backing layer to quench the resonance effect (fig. S6; see materials and methods for details on fabrication and optimization of the matching and backing layers). The BAUS probe is sealed by a layer of epoxy for high stability and reliability in long-term applications. We further designed a plug-and-play input-output to connect with the flexible flat cable (fig. S7). The BAUS probe has a thickness of 3 mm, length and width of 1 to 2 cm, and weight of 10 to 40 g (fig. S3 and table S3).

The BAUS couplant consists of a soft yet tough hydrogel composed of chitosan-polyacrylamide interpenetrating polymer networks (10 wt %) and water (90 wt %) (see materials and methods for details on fabrication of the BAUS couplant). The hydrogel is encapsulated by a thin elastomer membrane (thickness of <40 μm) of polyurethane to prevent dehydration of the hydrogel and to provide comfortable skin contact with a dry couplant surface (figs. S8 and S9). The polyurethane membrane is grafted with poly(acrylic acid) coupled with *N*-hydroxysuccinimide ester (NHS ester) to form robust bonding with the hydrogel (*26*). The hydrogel-elastomer hybrid is further coated by a thin bioadhesive layer (thickness of <10 μm) synthesized by copolymerizing poly(ethylene glycol) diacrylate, 2-ethylhexyl acrylate, acrylic acid, and acrylic acid–NHS ester (fig. S10). The carboxylic acid, ethyl, and hexyl groups in the bioadhesive can form physical bonds such as hydrogen bonds and electrostatic interactions with the skin; the NHS ester groups in the bioadhesive can form covalent amide bonds with the skin (*26*). The total thickness of the elastomer membrane and the bioadhesive layer is less than one-fourth of the acoustic wavelength to minimize the acoustic attenuation of the BAUS couplant (Fig. 2D).

To quantitatively validate that the BAUS device can robustly adhere to the skin under

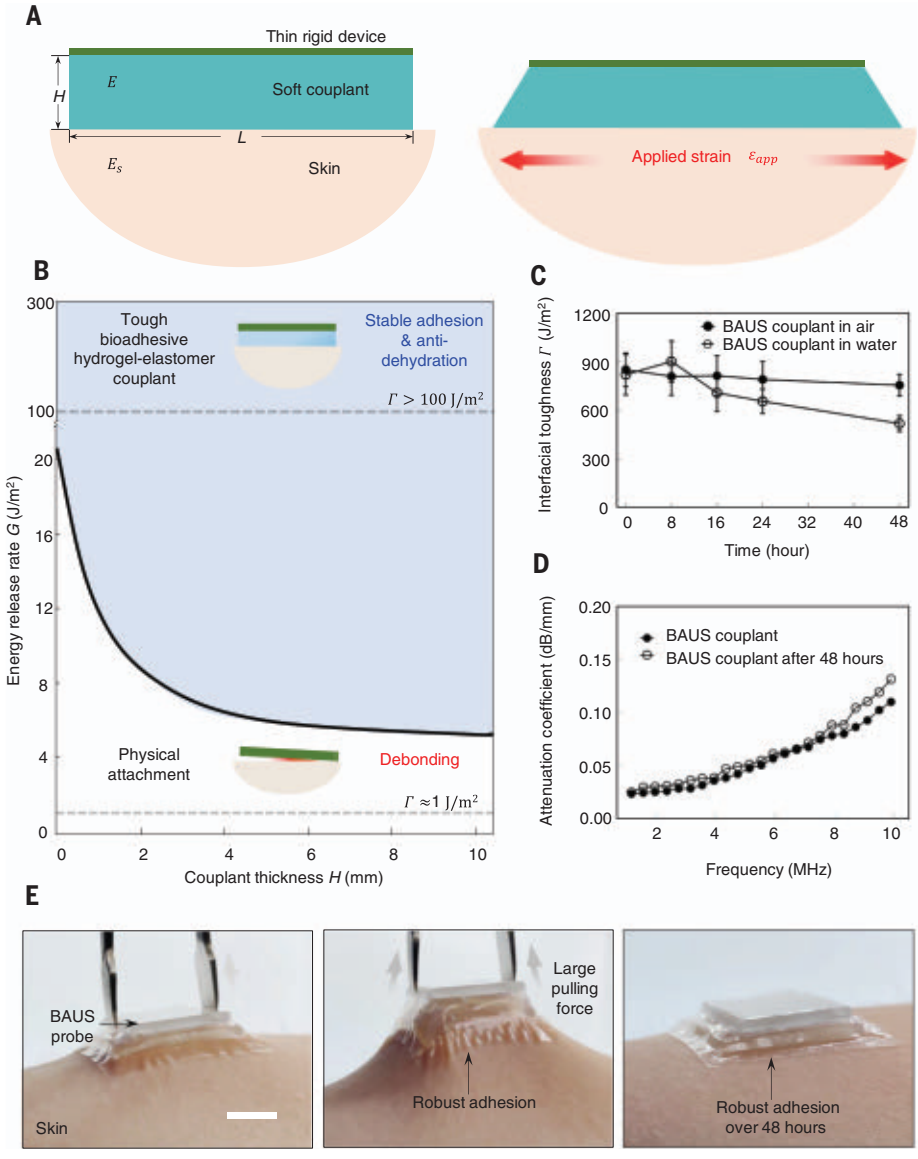


Fig. 2. Adhesion and acoustic performances of the BAUS device. (**A**) Schematics of the mechanical model for a rigid device adhered on the skin via a couplant layer. E, couplant Young's modulus; E_s, skin Young's modulus; H, couplant thickness; L, device length. (**B**) The energy release rate G as a function of the couplant thickness H calculated by the model with typical parameters: skin and couplant Young's modulus $E_s = E =$ 100 kPa, device length $L = 2$ cm, applied strain on the skin $\varepsilon_{app} = 20\%$, and interfacial toughness between the couplant and the skin Γ. (**C**) Interfacial toughness between the BAUS couplant and porcine skin measured over 48 hours after adhesion formation. Error bars indicate standard deviation (±SD) of interfacial toughness ($n = 4$ independent samples). (**D**) Acoustic attenuation coefficients of the BAUS couplant at frequencies from 1 to 10 MHz over 48 hours in air. (**E**) The BAUS device adhered on the skin can withstand high pulling forces and maintain robust adhesion on the skin over 48 hours. Scale bar is 10 mm.

dynamic body motions, we developed a model of a rigid device adhered to the skin via a soft couplant layer (Fig. 2A; see fig. S11 and materials and methods for details on the model). When the skin is stretched under an applied strain ε_{app}, a crack may initiate at the edge of the couplant-skin interface and propagate. Crack propagation is driven by the energy release rate G (defined as the reduction of stored elastic energy in the couplant-skin system when the crack propagates by a unit area) and resisted by the interfacial toughness Γ (defined as the energy required to propagate the crack by a unit area). In Fig. 2B, we calculate the energy release rate G of a couplant-skin system under the plane-strain condition with a typical device length L of 2 cm and a typical applied strain on the skin ε_{app} of 20%. For simplicity of calculation, the skin and the couplant layer are taken as incompressible linear elastic materials with the same Young's modulus of 100 kPa. If the energy release rate is higher than the interfacial toughness $G > \Gamma$ (unshaded zone in Fig. 2B), the crack propagates on the interface and the device debonds from the skin. When the rigid device is physically attached on the skin without any couplant (couplant thickness $H = 0$), the interfacial toughness is ~1 J/m^2, and the rigid device detaches from the skin. By contrast, when the rigid device is adhered to the skin via the BAUS couplant, the interfacial toughness exceeds 500 J/m^2 both in air and under water (Fig. 2C), robustly bonding the device on the skin for any couplant thickness. In fig. S11, we calculate the energy release rate of the couplant-skin system with various parameters for the future design of bioadhesive devices.

To evaluate the adhesion performance of the BAUS couplant on the skin, we conducted the standard 90°-peeling test (ASTM D2861) to measure the interfacial toughness of the elastomer membrane adhered on porcine skin via the bioadhesive layer (see materials and methods for details on the test). We choose porcine skin as the model tissue for evaluating adhesion performance because of its resemblance to human skin (*26*). The elastomer membrane can establish tough adhesion on porcine skin upon contact, providing an interfacial toughness of 853 J/m^2. When the adhered sample is stored in air (relative humidity of 30 to 50%; temperature of 24°C), the interfacial toughness decreases to 756 J/m^2 over 48 hours. When the adhered sample is stored in water to simulate wet environments, the interfacial toughness decreases to 518 J/m^2 over 48 hours. These results (Fig. 2, C and E) show that the bioadhesive layer can robustly adhere the BAUS device on the skin over 48 hours in dry and wet environments. The BAUS device can be detached from the skin by peeling from one edge of the BAUS couplant without leaving any residue (fig. S12). In comparison, the interfacial toughness of typical liquid (Aquasonic Clear, Parker Labs) and solid (Aquaflex, Parker Labs) hydrogel couplants and elastomer couplants, including silicone (Ecoflex, Smooth-On), polyurethane (Sigma-Aldrich), and Aqualene (Olympus), are all less than 10 J/m^2, which cannot give stable adhesion of ultrasound probes on the skin. Although the 3M VHB acrylic tape can maintain a relatively high interfacial toughness (>456 J/m^2) in air over 48 hours, the interfacial toughness reduces to 27 J/m^2 in water, possibly leading to unstable adhesion in wet environments.

To evaluate the wearing comfort of the BAUS couplant on the skin, we adhered a set of BAUS couplants on various body locations—including arm, neck, chest, and waist—of 15 subjects for 48 hours (fig. S9). Most of the subjects reported no positive reaction (i.e., itchiness, irritation, dryness, or redness) at any

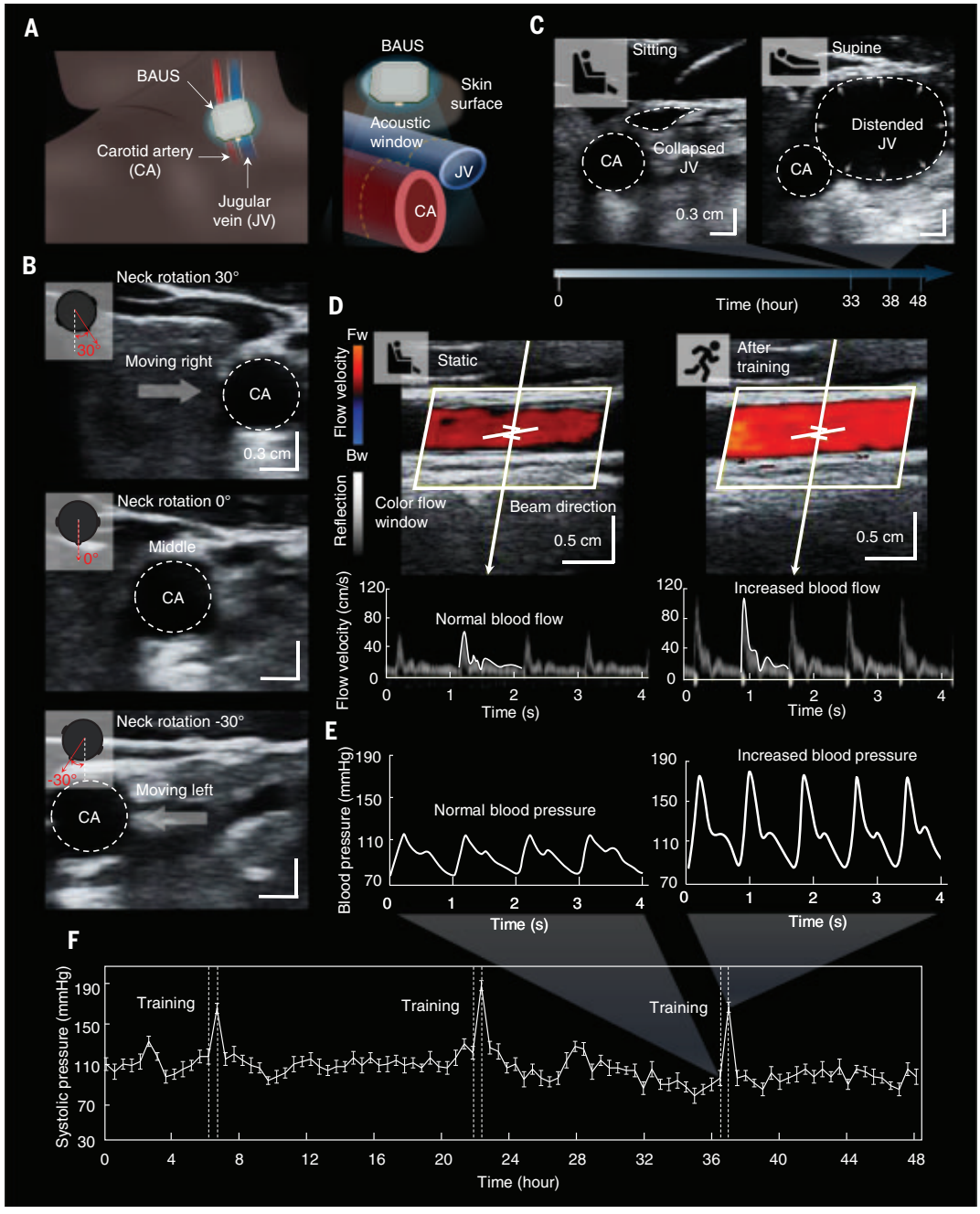


Fig. 3. Long-term continuous BAUS imaging of blood vessels. See also fig. S30 and movie S5. (**A**) Schematics of the BAUS device adhered on the neck of the subject to image the carotid artery (CA) and jugular vein (JV). (**B**) The BAUS device can stably image the blood vessels under dynamic body motions such as neck rotation with angles up to ±30°. (**C**) The diameter of the jugular vein increases substantially from the sitting or standing position to the supine position. (**D**) Color-flow imaging of the carotid artery by the BAUS device. The diameter of the carotid artery and the blood flow rate in the carotid artery increase substantially after 0.5 hours of physical exercise. Fw, forward; Bw, backward. (**E**) Blood pressure waveforms of the carotid artery before and after 0.5 hours of physical exercise. The systolic blood pressure of the carotid artery increases substantially after 0.5 hours of physical exercise. The decrease in the blood pressure from the systolic peak in each cardiac cycle is much steeper after 0.5 hours of physical exercise. (**F**) The systolic blood pressure of the carotid artery over 48 hours as measured from the BAUS imaging. The systolic blood pressure is measured by calculating the mean and standard deviation of 15 consecutive systolic peaks every 0.5 hours. Error bars indicate standard deviation (±SD) of measured systolic pressure (n = 15 independent samples). The 0.5-hour training periods are indicated by the dashed lines.

location; only one subject reported minor itchiness after 48 hours of continuous wearing (fig. S13). As control samples, the abovementioned typical couplants were adhered (3M VHB acrylic tape) or mechanically wrapped (others) on the arms of the 15 subjects. The liquid and solid hydrogel couplants received the highest positive reactions, and the positive reactions of the elastomer couplants were also higher than that of the BAUS couplant. These results are consistent with the comfort level of the subjects as evaluated by the visual analog scale (VAS): the BAUS couplant is most comfortable, and the liquid and solid hydrogel couplants are least comfortable (fig. S13).

To evaluate the acoustic performance of the BAUS couplant, we used the transmission-through method to measure the attenuation coefficients of the BAUS couplant. At frequencies from 1 to 10 MHz, the attenuation coefficients of the BAUS couplant were measured to be from 0.036 to 0.131 dB/mm (Fig. 2D and fig. S14). We further measured the acoustic impedance of the BAUS couplant to be 1.59 MRayl, which matches the acoustic impedance of the skin. In comparison, the typical liquid and solid hydrogel couplants have similar attenuation coefficients as those of the BAUS couplant at frequencies from 1 to 10 MHz: 0.070 to 0.120 dB/mm and 0.117 to 0.174 dB/mm, respectively. However, the attenuation coefficients of all abovementioned elastomer couplants are multiple times higher than that of the BAUS couplant at the same frequency. This is consistent with the result that these elastomer couplants are too damping to allow wearable ultrasound imaging of internal organs (fig. S2).

To evaluate the performances of the BAUS probes, we first performed electrical impedance measurements on all elements in the probes (fig. S15 and table S4). The measured center frequency of the elements on each BAUS probe is indeed 3, 7, or 10 MHz (with a minor deviation of <5%) as prescribed; the electrical impedances of the probes at their center

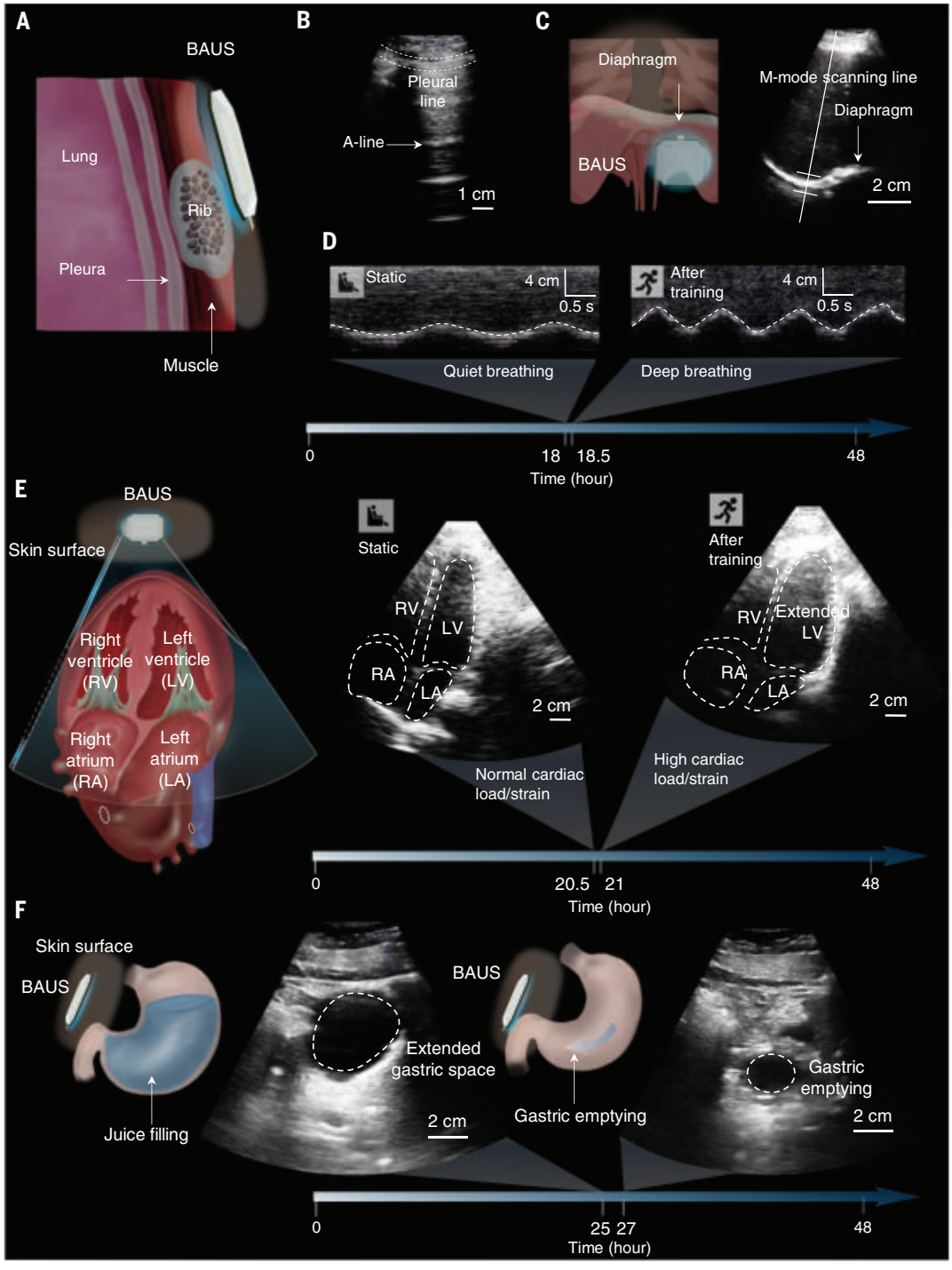


Fig. 4. Long-term continuous BAUS imaging of lung, diaphragm, heart, and stomach. See also figs. S33 to S36. (**A**) Schematics of the BAUS device adhered to the chest of the subject to image the lung. (**B**) The BAUS imaging shows a smooth pleural line and repetitive A-lines of the lung. (**C**) The BAUS imaging (right) shows clear diaphragm morphology. A diagram of the placement of the BAUS device is shown on the left. (**D**) The diaphragm motility increases substantially in amplitude and frequency after 0.5 hours of physical exercise. (**E**) The BAUS imaging shows the dynamics of the four chambers of the heart. The size of the left ventricle greatly increases after 0.5 hours of physical exercise. (**F**) The BAUS imaging shows the dynamics of the stomach. The gastric antral cross-sectional area gradually decreases after the subject drinks 450 ml of juice.

frequencies are 49, 36, or 43 ohms, respectively (fig. S16 and table S4). We then measured the cross-talk in each BAUS probe to be less than −40 dB, which is sufficiently low for high-quality imaging applications (fig. S17 and table S4) (*27*). We next performed the pulse-echo tests on all BAUS probes to measure their bandwidths. The measured bandwidths of the 3-, 7-, and 10-MHz BAUS probes are 68, 75, and 78%, respectively (fig. S18 and table S4).

To evaluate the imaging resolutions of the BAUS probes, we first performed the pulse-echo tests on the BAUS probes to characterize their axial resolutions (fig. S19). The axial resolutions of the 3-, 7-, and 10-MHz BAUS probes are 0.77, 0.225, and 0.1924 mm, respectively (fig. S19 and table S4). We then measured the point spread functions of the BAUS probes to characterize their lateral resolutions. The lateral resolutions of the 3-, 7-, and 10-MHz BAUS probes are 1.79, 0.38, and 0.38 mm at their imaging focal depths of 6, 3, and 2 cm, respectively (figs. S20 to S22 and table S4). These measured values are also confirmed by the acoustic simulations (figs. S20 to S22). In comparison, the axial resolution of the stretchable ultrasound imaging probe (2 MHz) is reported to be 5.775 mm (*12*).

To test the imaging stability of the BAUS device, we adhered the BAUS device on a

phantom of a blood vessel in soft tissues under various frequencies of vibration (0 to 3.5 Hz with an amplitude of 2 cm) to evaluate its acoustic performance under dynamic body motions. Because of stable adhesion of the BAUS device to the phantom and the relatively high imaging frame rate (40 Hz), the motion artifact on imaging is minimal (fig. S23).

To test the thermal properties of the BAUS device, we measured the temperature of the BAUS device when continuously actuated by various levels of working voltages (0 to 90 V) over 48 hours. Owing to its low working power and thin form factor for heat dissipation, the BAUS device can maintain a constant level of temperature (24°C) while continuously imaging for 48 hours (figs. S24 and S25), validating its thermal stability in long-term applications.

We demonstrate the BAUS imaging platform's capability to perform 48-hour continuous imaging of human organs, including blood vessels and heart, muscle and diaphragm, stomach, and lung. These organs represent the cardiovascular, muscular, digestive, and respiratory systems, respectively. To image organs with depths less than 6 cm beneath the skin (i.e., jugular vein, carotid artery, and bicep muscle), we chose an ultrasound probe with a center frequency of 7 or 10 MHz and adopted the plane wave compounding method (fig. S26). To image organs deeper than 6 cm beneath the skin (i.e., heart, stomach, diaphragm, and lung), we chose an ultrasound probe with a center frequency of 3 MHz and adopted the phased array imaging method (figs. S27 to S29 and movies S3 and S4; see materials and methods for details on acoustic simulation). The probes are adhered to various locations on the skin of a healthy subject via the BAUS couplants for continuous imaging of the corresponding organs over 48 hours (see materials and methods for details on data acquisition and image processing).

Figure 3 and fig. S30 demonstrate the imaging results of the jugular vein and carotid artery obtained by the 10-MHz BAUS device adhered to the right-hand side of the neck (Fig. 3A). Dynamic body motions such as neck rotation can cause blood vessel shifting and skin deformation, which deteriorate the imaging quality and stability of existing wearable ultrasound devices (*11*). By contrast, the BAUS device can provide 48-hour continuous imaging of the jugular vein and carotid artery under dynamic body motions such as neck rotation with angles up to ±30° (Fig. 3B and movie S5). This is because of the robust adhesion of the BAUS device on the neck, the high imaging resolution of the BAUS probe, and the large acoustic window given by the BAUS probe aperture (2 cm by 1 cm). The BAUS imaging results show that the diameter of the jugular vein in the subject increases from the sitting or standing position to the supine position (Fig. 3C and fig. S30). The change of the jugular vein diameter strongly correlates with the right atrium pressure (*28*) and thus can be used to diagnose cardiac diseases such as heart failure and pulmonary hypertension (*29*). The BAUS imaging results further provide 48-hour continuous blood flow rate and blood pressure waveform data of the carotid artery (Fig. 3, D to F, and fig. S30). For example, we observed dramatic increases in blood flow rate (from 65 to 117 cm/s; Fig. 3D) and systolic blood pressure (from 115 to 168 mmHg; Fig. 3E) of the carotid artery in the subject before and after 0.5 hours of physical exercise (see materials and methods for details on data acquisition and calculation of the blood pressure). Moreover, from the blood pressure waveform measured after the 0.5-hour jogging session (right panel of Fig. 3E), we can see a steeper decrease in the blood pressure from the systolic peak in each cardiac cycle than before the exercise, possibly because of the exercise-induced vasodilation of the carotid artery.

Figure S31 demonstrates the imaging results of the bicep muscle obtained by the 7-MHz BAUS device adhered to the upper arm of the subject (fig. S31). The BAUS system can provide 48-hour continuous imaging of the bicipital muscle. For example, we observed increased blood perfusion and echogenicity on the muscle and tendon interface 2 hours after a 1-hour weightlifting training session and increased vascularity 10 hours after the training (fig. S32).

Figure 4, A and B, and fig. S33 demonstrate the imaging results of the lung obtained by the 3-MHz BAUS phased-array device adhered on the right chest (longitudinal direction) of the subject (Fig. 4A). The BAUS system can provide 48-hour continuous imaging of key characteristics of the lung, including the pleural line and A-lines. The BAUS device is also able to continuously and stably image the lung under body motions such as jogging and cycling (fig. S34). The BAUS imaging shows a smooth pleural line and repetitive A-lines of the subject over 48 hours, indicating a healthy, normally aerated lung. Although lung ultrasonography has proven to be an effective tool for diagnosing COVID-19 patients (*30*), it is mostly available in hospitals for infrequent and noncontinuous imaging of the lung. By contrast, the wearable BAUS system could continuously monitor the symptoms of possibly infected COVID-19 patients at home.

Figure 4, C and D, demonstrates the imaging results of the diaphragm motility obtained by the 3-MHz BAUS phased-array device adhered on the right anterior subcostal region of the subject before and after 0.5 hours of physical exercise. The BAUS imaging shows that the diaphragm motility after training is more dramatic in amplitude and frequency compared with quiet breathing before training.

Figure 4E and fig. S35 demonstrate the imaging results of the heart obtained by the 3-MHz BAUS phased-array device adhered on the left chest of the subject (as visualized through the apical four-chamber view). The BAUS system can provide 48-hour continuous imaging of the dynamics of the four cardiac chambers: the right ventricle, right atrium, left ventricle, and left atrium. The BAUS device is also able to continuously and stably image the heart under body motions such as jogging and cycling (fig. S34). For example, we observed that the size of the left ventricle greatly increases after 0.5 hours of physical exercise. The continuous images of the cardiac chambers (fig. S35) can be further processed to calculate the dynamics of cardiac strain, which is an important parameter that could indicate cardiomyopathy.

Figure 4F and fig. S36 demonstrate the imaging results of the stomach obtained by the 3-MHz BAUS phased-array device adhered on the upper abdomen skin of the subject. The BAUS imaging system provides 48-hour continuous imaging of the gastric antral cross section. For example, we observed an extended gastric antrum once the subject drank 450 ml of juice (Fig. 4F). Over the next 2 hours, the gastric antral cross-sectional area gradually decreased because of gastric emptying (Fig. 4F and fig. S36).

Although the imaging resolutions of the BAUS probes are superior to those of existing wearable ultrasound devices (Fig. 1G and table S1), the BAUS probes can be further improved for better imaging quality in the future. The pitch of the current BAUS probes is relatively large (0.5 mm; table S3) and can cause grating lobes, especially in the 7- and 10-MHz BAUS probes. Although the grating-lobe artifacts can be alleviated by using the plane wave compounding method (fig. S26) (*31*), the pitches of future BAUS probes will be reduced to eliminate the artifacts. Although the current BAUS probes do not have elevation focusing, thin-profile (<1 mm) elevation lenses (*32*) can be added to future BAUS probes for elevation focusing to achieve better imaging quality, especially for deep organs.

Because the current work is focused on designing and fabricating the BAUS devices, we use an external system (Verasonics Vantage system) for data acquisition. Despite the large size of the data acquisition system, the BAUS devices may find immediate applications in nonmobile imaging in clinical settings such as long-term imaging of patients in intensive-care units. In addition, because the data acquisition systems of point-of-care ultrasound devices (e.g., GE VScan, Butterfly IQ, and Phillips Lumify) have been miniaturized to the size of a cell phone, the data acquisition system of BAUS can be miniaturized in the future. Beyond clinical applications, the BAUS imaging platform could provide long-term continuous imaging of the developments of tissues and

organs—including embryo, tumors, and brain—in a noninvasive manner (*33*, *34*).

The current paradigm for the biointegration of devices is to make the devices thin and stretchable for conformal attachment on the body. Here, we propose a different paradigm for biointegration: to robustly adhere thin and rigid devices on the body via a soft, tough, and bioadhesive couplant (Figs. 1 and 2). Developing bioadhesive couplants that enable electrical, optical, thermal, chemical, and/or biological interfacing with the body could open new avenues to future biointegrated devices.

REFERENCES AND NOTES

1. E. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Basic Books, 2019).
2. M. Lin, H. Hu, S. Zhou, S. Xu, *Nat. Rev. Mater.* (2022).
3. D.-H. Kim *et al*., *Science* **333**, 838–843 (2011).
4. S. Xu *et al*., *Science* **344**, 70–74 (2014).
5. S. Yang *et al*., *Adv. Mater.* **27**, 6423–6430 (2015).
6. W. Gao *et al*., *Nature* **529**, 509–514 (2016).
7. H. Lee *et al*., *Nat. Nanotechnol.* **11**, 566–572 (2016).
8. J. R. Sempionatto *et al*., *Nat. Biomed. Eng.* **5**, 737–748 (2021).
9. C. L. Moore, J. A. Copel, *N. Engl. J. Med.* **364**, 749–757 (2011).
10. C. Wang *et al*., *Nat. Biomed. Eng.* **2**, 687–695 (2018).
11. S. R. Steinhubl, E. J. Topol, *Nat. Biomed. Eng.* **2**, 633–634 (2018).
12. C. Wang *et al*., *Nat. Biomed. Eng.* **5**, 749–758 (2021).
13. C. K. L. Phoon, *Pediatr. Res.* **60**, 14–21 (2006).
14. Q. Huang, J. Lan, X. Li, *IEEE Trans. Industr. Inform.* **15**, 1173–1182 (2019).
15. H. T. Sen, M. A. L. Bell, I. Iordachita, J. Wong, P. Kazanzides, in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2013), pp. 3071–3076.
16. H. M. Heres *et al*., *J. Med. Ultrason.* **47**, 47–56 (2020).
17. R. W. Nuckols *et al*., *Sci. Robot.* **6**, eabj1362 (2021).
18. M. Sjoerdsma *et al*., *Appl. Sci.* **11**, 1534 (2021).
19. H. R. Laine, A. Harjula, P. Peltokallio, *J. Ultrasound Med.* **6**, 33–36 (1987).
20. F. Chasset *et al*., *J. Dermatol.* **43**, 318–320 (2016).
21. E. Ginzel, R. MacNeil, R. Ginzel, M. Zuber, A. Sinclar, *NDT.net J.* **20**, 1–12 (2015).
22. J. S. Kenny *et al*., *Sci. Rep.* **11**, 7780 (2021).
23. B. Y. Ahn *et al*., *Science* **323**, 1590–1593 (2009).
24. Z. Huang *et al*., *Nat. Electron.* **1**, 473–480 (2018).
25. G. Hong *et al*., *Science* **360**, 1447–1451 (2018).
26. H. Yuk *et al*., *Nature* **575**, 169–174 (2019).
27. Q. Zhou, K. H. Lam, H. Zheng, W. Qiu, K. K. Shung, *Prog. Mater. Sci.* **66**, 87–111 (2014).
28. S. Magder, *Crit. Care Med.* **34**, 2224–2227 (2006).
29. A. E. Raine *et al*., *N. Engl. J. Med.* **315**, 533–537 (1986).
30. C. Xing *et al*., *Crit. Care* **24**, 174 (2020).
31. M. Tanter, M. Fink, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **61**, 102–119 (2014).
32. D. G. Wildes *et al*., *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44**, 1027–1037 (1997).
33. L. V. Wang, S. Hu, *Science* **335**, 1458–1462 (2012).
34. C. Errico *et al*., *Nature* **527**, 499–502 (2015).
35. C. Wang, X. Chen, L. Wang, M. Makihata, H.-C. Liu, T. Zhou, X. Zhao, Bioadhesive ultrasound for long-term continuous imaging of diverse organs, Version 1. Harvard Dataverse (2022).

ACKNOWLEDGMENTS

We thank Z. Zhang and B. Anthony for technical support on imaging algorithms, R. Chen for advice on device packaging, S. Lin for help on interfacial toughness tests, Y. Cao for SEM sample preparation, and H. Roh for initial involvement in device fabrication. **Funding:** This work was supported by the Defense Advanced Research Projects Agency through cooperative agreement D20AC00004, the National Science Foundation (no. EFMA-1935291), the National Institutes of Health (no. 1R01HL153857), and the US Army Research Office through the Institute for Soldier Nanotechnologies at MIT (no. W911NF-13-D-0001). X.Z. acknowledges support from a George N. Hatsopoulos (1949) Faculty Fellowship and a Humbolt Research Award. **Author contributions:** C.W. and X.Z. conceived the project. X.C. and X.Z. designed the BAUS couplant, and X.C. fabricated and characterized the BAUS couplant. T.Z. took SEM images of the BAUS couplant. C.W. designed the BAUS probe, and C.W. and X.C. fabricated and characterized the BAUS probe. H.-C.L. designed the imaging algorithms. M.M. designed the circuit adaptor. C.W. performed the acoustic simulation. L.W. performed the mechanical simulation. C.W. and X.C. tested the BAUS device in vitro and in vivo and analyzed the results. C.W., X.Z., X.C., and L.W. wrote the paper and incorporated comments and edits from all authors. X.Z. supervised the project. **Competing interests:** X.Z., C.W., and X.C. are inventors on a patent application describing the BAUS system and technology. X.Z. has a financial interest in SanaHeal, Inc., a biotechnology company focused on the development of medical devices for surgical sealing and repair. The authors declare that they have no other competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The website associated with the manuscript, including data and algorithms, is available at Harvard Dataverse (*35*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo2542
Materials and Methods
Figs. S1 to S36
Tables S1 to S4
References (*36*, *37*)
Movies S1 to S5

Submitted 22 January 2022; resubmitted 4 April 2022
Accepted 9 June 2022
10.1126/science.abo2542

COASTAL ECOLOGY

Constraints on the adjustment of tidal marshes to accelerating sea level rise

Neil Saintilan[1]*, Katya E. Kovalenko[2], Glenn Guntenspergen[3], Kerrylee Rogers[4], James C. Lynch[5], Donald R. Cahoon[3], Catherine E. Lovelock[6], Daniel A. Friess[7], Erica Ashe[8], Ken W. Krauss[9], Nicole Cormier[1], Tom Spencer[10], Janine Adams[11], Jacqueline Raw[11], Carles Ibanez[12], Francesco Scarton[13], Stijn Temmerman[14], Patrick Meire[15], Tom Maris[15], Karen Thorne[16], John Brazner[15], Gail L. Chmura[17], Tony Bowron[18], Vishmie P. Gamage[1], Kimberly Cressman[19], Charlie Endris[20], Christina Marconi[21], Pamela Marcum[22], Kari St. Laurent[23], William Reay[24], Kenneth B. Raposa[25], Jason A. Garwood[26], Nicole Khan[27]

Much uncertainty exists about the vulnerability of valuable tidal marsh ecosystems to relative sea level rise. Previous assessments of resilience to sea level rise, to which marshes can adjust by sediment accretion and elevation gain, revealed contrasting results, depending on contemporary or Holocene geological data. By analyzing globally distributed contemporary data, we found that marsh sediment accretion increases in parity with sea level rise, seemingly confirming previously claimed marsh resilience. However, subsidence of the substrate shows a nonlinear increase with accretion. As a result, marsh elevation gain is constrained in relation to sea level rise, and deficits emerge that are consistent with Holocene observations of tidal marsh vulnerability.

Tidal marshes are among the most vulnerable of the world's ecosystems. Throughout human civilization, tidal marshes have been reclaimed for agriculture and settlement, and the pace of loss has accelerated in concert with burgeoning coastal populations on all inhabited continents over the past century (*1*, *2*). To this pressure has been added the threat of accelerating sea level rise. Because tidal marshes occur within tightly defined elevation ranges relative to mean sea level, they are sentinel ecosystems at the forefront of coastal climate change impact. Potential tidal marsh loss with sea level rise threatens a range of ecosystem services valued at US ~$27 trillion per year (*3*), extending to fisheries production, recreation, cultural heritage, coastal protection, water quality enhancement, and carbon sequestration.

Sea level rise can lead to marsh loss through marsh edge erosion, conversion to mudflats, encroachment of mangrove forests where they

[1]School of Natural Sciences, Macquarie University, Sydney, NSW, Australia. [2]Natural Resources Research Institute, University of Minnesota, Duluth, MN, USA. [3]US Geological Survey, Eastern Ecological Science Center, Beltsville, MD, USA. [4]School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, NSW, Australia. [5]National Park Service, Washington, DC, USA. [6]School of Biological Sciences, University of Queensland, Brisbane, QLD, Australia. [7]Department of Geography, National University of Singapore, Singapore. [8]Department of Earth and Planetary Sciences, Rutgers University, Newark, NJ, USA. [9]US Geological Survey, Wetland and Aquatic Research Center, Lafayette, LA, USA. [10]Cambridge Coastal Research Unit, Department of Geography, Cambridge University, Cambridge, UK. [11]Institute for Coastal and Marine Research and Department of Botany, Nelson Mandela University, Gqeberha, South Africa. [12]Eurecat, Unit of Climate Change, Centre Tecnològic de Catalunya, Catalonia, Spain. [13]SELC Societa Cooperativa, Venice, Italy. [14]Ecosphere Group, University of Antwerp, Antwerp, Belgium. [15]Nova Scotia Department of Natural Resources and Renewables, Nova Scotia, Canada. [16]US Geological Survey, Western Ecological Research Center, Davis, CA, USA. [17]Department of Geography, McGill University, Montreal, Canada. [18]Saint Mary's University, Halifax, Canada. [19]Mississippi State University, Starkville, MS, USA. [20]Moss Landing Marine Labs, California State University, Moss Landing, CA, USA. [21]Marine Science Institute, University of Texas, Austin, TX, USA. [22]Guana Tolomato Matanzas National Estuarine Research Reserve, Ponte Vedra Beach, FL, USA. [23]Delaware Department of Natural Resources and Environmental Control, Dover, DE, USA. [24]Virginia Institute of Marine Science, Gloucester Point, VA, USA. [25]Narragansett Bay National Estuarine Research Reserve, Prudence Island, RI, USA. [26]Apalachicola National Estuarine Research Reserve, Eastpoint, FL, USA. [27]Department of Earth Sciences, Swire Institute of Marine Science, University of Hong Kong, Hong Kong, China.
***Corresponding author. Email: neil.saintilan@mq.edu.au**

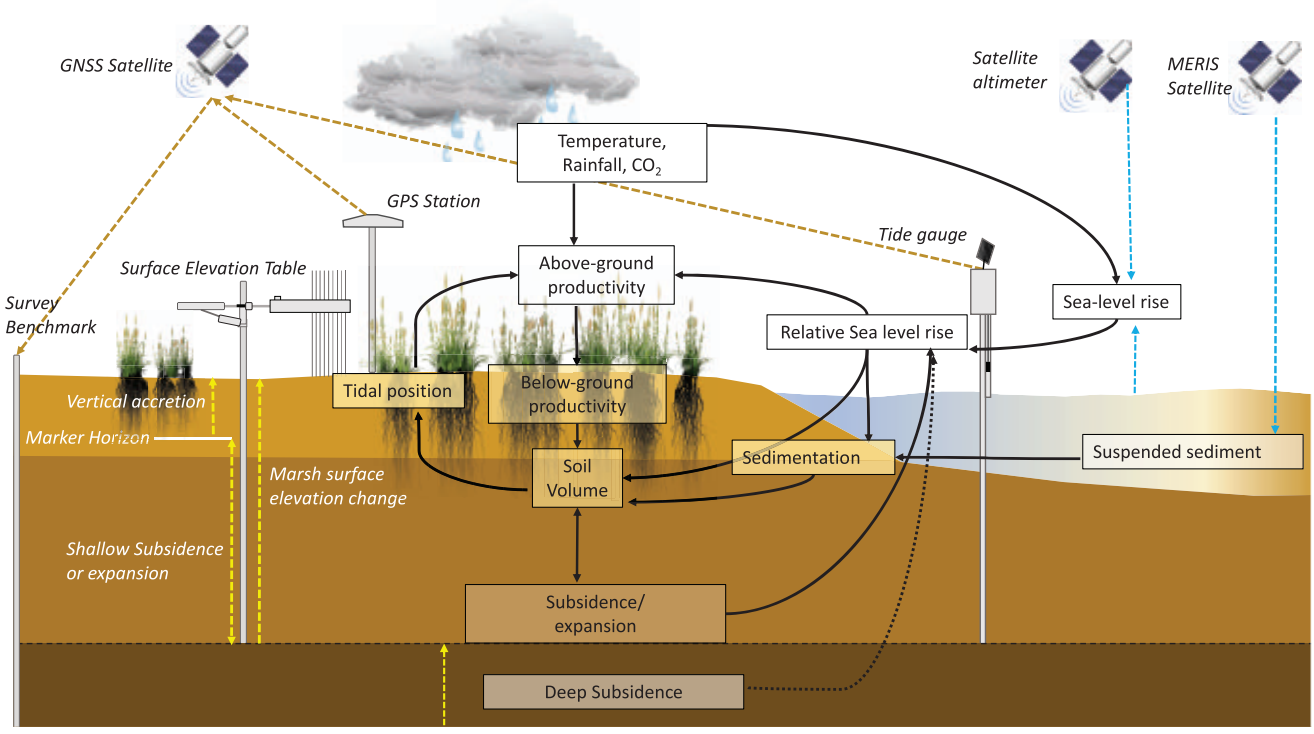


Fig. 1. Processes influencing marsh surface elevation and their measurement using the surface elevation table–marker horizon (SET-MH) monitoring station. Feedbacks among sea level rise, vertical accretion, and elevation gain are conceptualized as driving marsh substrates toward an equilibrium elevation within the tidal frame, facilitated by inputs of mineral and organic matter. The SET-MH method measures soil elevation relative to a benchmark (to which the portable component of the SET is attached), while a tide gauge records the combined effect of changes in sea level and land movement occurring below the survey benchmark rod, to which the gauge is routinely leveled. This combined recording of eustatic sea level (ocean volume) change and deep land movement is termed relative sea level rise (RSLR) and does not include processes measured by the SET-MH method. GNSS, Global Navigation Satellite System; MERIS, Medium Resolution Imaging Spectrometer.

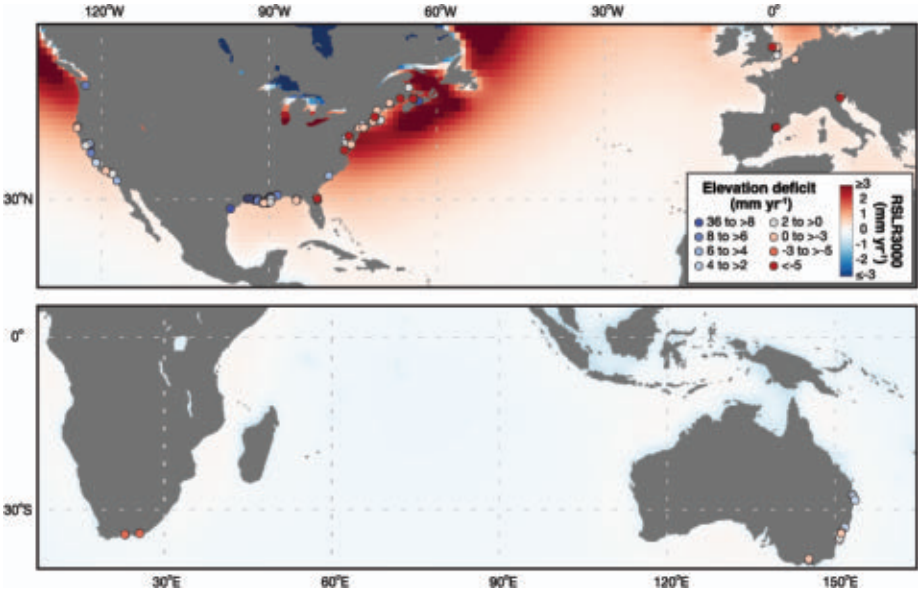


Fig. 2. Distribution of tidal marsh SET-MH monitoring stations used in the analyses, and deficit between elevation gain and contemporaneous local RSLR. Deficits are assigned positive numbers. The background of late Holocene (0 to 3000 years B.P.) RSLR is derived from glacio-isostatic modeling (*20*).

occupy lower tidal position, and/or the expansion of internal ponds and channels, with all mechanisms enhanced by the loss of marsh surface elevation relative to mean tide level (*4*). The fate of tidal marshes under accelerating sea level rise will be determined not only by opportunities for landward marsh migration (*5*) but also by the capacity of tidal marshes to gain elevation through vertical accretion of mineral sediment and organic matter (*6*). Biophysical feedbacks between sea level rise and the vertical development of marsh substrates reduce the risk of loss occurring through conversion to unvegetated mudflats or open water (*7*).

Modeling based on observations from US East Coast organic marshes (*8*) and UK minerogenic marshes (*9*) has suggested that an equilibrium may emerge among the position of a marsh within the tidal frame, plant productivity, root mass development, sedimentation, and the elevation of the marsh in response to mean sea level (Fig. 1). This equilibrium may be sustained under low rates of relative sea level rise (RSLR), the combination of vertical land movement and sea level change. How widely these controls and their upper limits operate across marsh sites globally has been a crucial and disputed question in the regional- to global-scale modeling of tidal marsh responses to projected rates of RSLR under climate change (*5*, *7*, *10*). At present, observations of contemporary marsh accretion suggest that marshes can adjust to rates of RSLR of >10 mm $year^{-1}$ (*7*, *11*, *12*). However, the Holocene marsh record suggests that adjustment is highly unlikely (90% probability) at RSLR exceeding 7 mm $year^{-1}$ for UK tidal marshes (*13*) and

tropical mangroves (*10*), and 3 to 5 mm year^{-1} for marshes in the Gulf of Mexico (*14*). Here, we report on contemporary tidal marsh elevation gain in relation to RSLR, testing the importance of environmental conditions in mediating these responses.

Several factors may influence the efficacy of tidal marsh vertical adjustment to sea level rise, but their relative contributions to explaining observed regional to global variability in marsh responses remain poorly elucidated. Globally, tidal range in marshes can vary from a few centimeters to 16 m, and this variability will influence the susceptibility of marshes to drowning, particularly where the tidal range is low relative to the projected RSLR (*12*, *15*). The position within the tidal frame influences the depth and duration of inundation and the deposition of sediment, but it also influences mineral and organic accretion responses of the marsh vegetation occurring at these specific positions (*8*, *9*). Tidal hydrodynamics and river discharge contribute to sediment delivery and accumulation (*15*), and these may be modified by flow control structures (*16*). Plant productivity is influenced by climate (precipitation and temperature), salinity, nutrients (*17*), atmospheric CO_2, and vegetation composition, which in turn influence soil organic carbon accumulation and decomposition. The rate of RSLR varies between coastlines and continents, and millennial-scale variability in RSLR may also control soil organic content, which may increase with sea level rise when conditions are favorable (*18*). Sampling at regional to global scales across hydrogeomorphic settings and biogeographic regions can clarify the relative importance of these factors and determine the consistency of feedbacks facilitating marsh adjustment to RSLR.

Accurate measurements of tidal marsh vertical adjustment in relation to sea level require a fixed benchmark against which elevation gain or loss can be measured. To this end, the surface elevation table–marker horizon (SET-MH) method has been developed as a global standard (*19*) for monitoring tidal marsh responses to RSLR (Fig. 1). A rod is driven into the marsh to form a stable benchmark against which elevation change can be measured. Vertical accretion (the surface accumulation of inorganic sediment, organic sediment, and living roots) is also measured at most sites above an artificial marker horizon (typically white feldspar, clay, or sand) introduced at the time of the first measurements against the benchmark (*20*). Data from SET-MH stations have informed models of marsh resilience to RSLR (*7*, *12*), global projections of tidal marsh and mangrove change in the coming century (*5*, *21*, *22*), and the influence of vertical accretion on carbon sequestration (*23*, *24*).

We analyzed tidal marsh elevation adjustment in relation to RSLR from SET-MH monitoring stations that met our criteria of emergent tidal marsh vegetation, sufficient length of record (>3 years), and exclusion from hydrological or experimental manipulation. The resulting network of 477 tidal marsh SET-MH stations, across 97 sites, forms clusters in regions with distinct hydrogeomorphic histories and tidal and biogeographic characteristics thought to be important to marsh resilience (Fig. 2). In general, Southern Hemisphere stations (Australia and South Africa) are in estuaries subject to millennia of stable or falling sea levels, with micro- to mesotidal marshes on high, stable intertidal platforms typically low in percentage of soil organic carbon (table S1 and data S1). North Atlantic coastlines (Bay of Fundy, Canada; UK; Belgium) are predominantly macrotidal, have been subject until recently to relatively stable (< ±0.5 mm year^{-1}) or falling sea levels over the past few thousand years, have low soil organic carbon content, and are situated adjacent to waters with high total suspended matter. Coastlines in the network that are subject to low rates of sea level rise over the past few millennia include two large river deltas with a microtidal regime (the Mississippi, USA; the Ebro, Spain) and microtidal to mesotidal barrier estuaries and embayments (Venice Lagoon, Southern California; the US Gulf of Mexico chenier plains and estuaries).

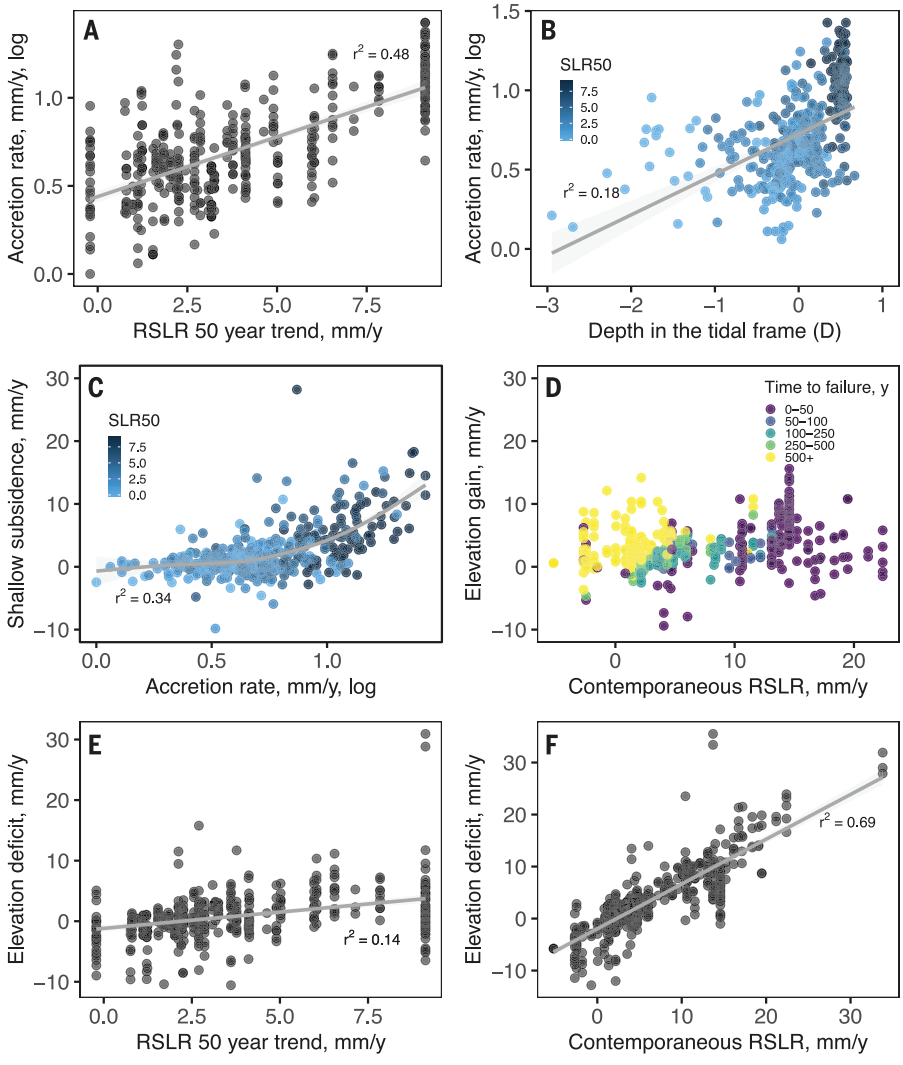


Fig. 3. The increasing vulnerability of tidal marshes to RSLR. (**A** and **B**) Accretion increases in parity with the 50-year RSLR trend (A) and with marshes lower in the tidal frame (B). (**C** and **D**) However, the rate of shallow marsh subsidence increases with the rate of vertical accretion, with an upward inflexion as RSLR increases between 5 and 10 mm year^{-1} (C), suppressing elevation adjustment to RSLR (D). (**E** and **F**) As a result, the deficit between elevation gain and RSLR increases with the 50-year RSLR trend (E) and the contemporaneous RSLR trend, the period over which individual SET-MH stations were measured (F). In (B) and (C), points are colored for the 50-year RSLR trend in mm year^{-1}, and in (D) for estimated time to failure (years) under the elevation deficit against the 50-year RSLR trend (*20*).

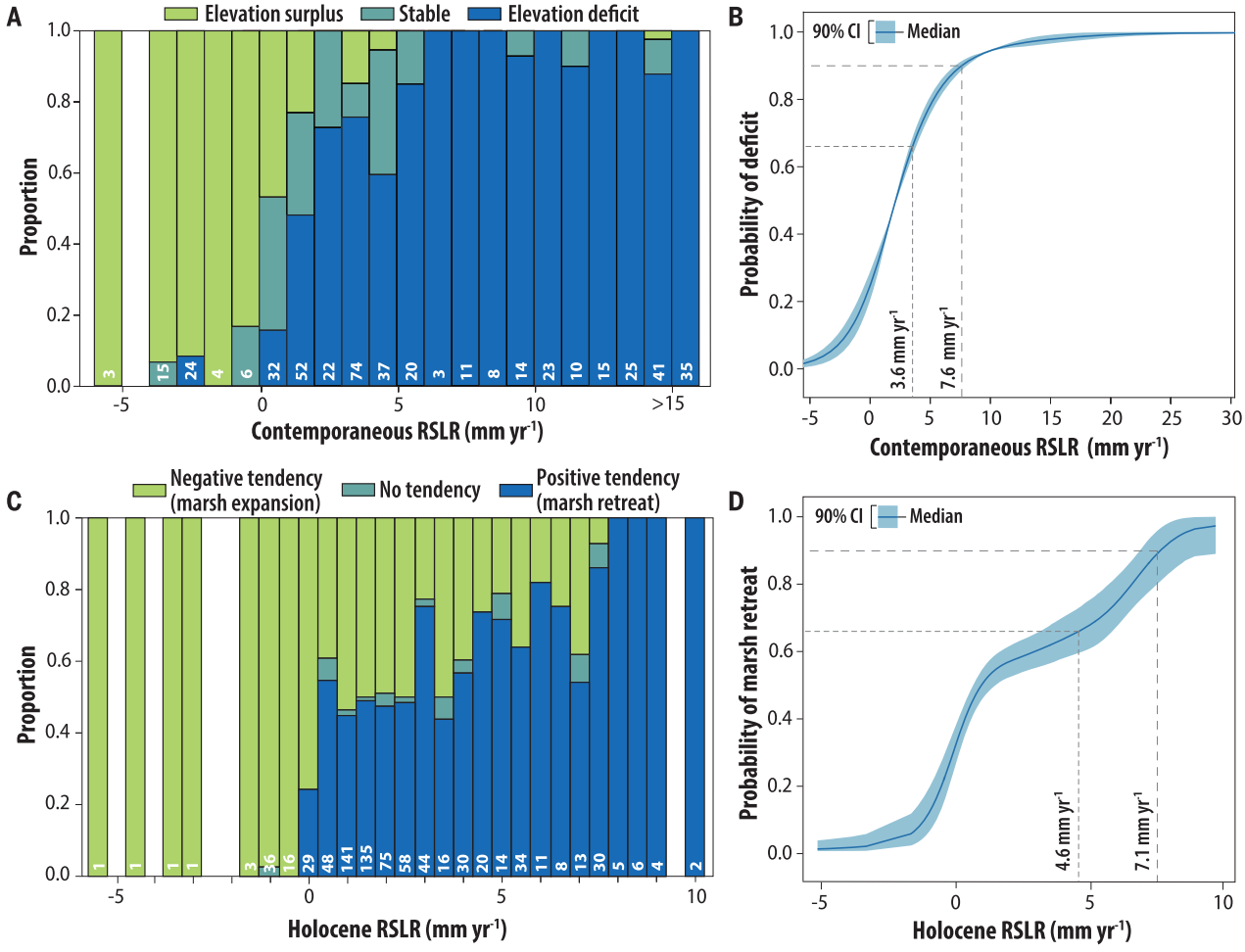


Fig. 4. Rates of relative sea level rise and marsh responses in the observational and paleo record. (**A**) Histogram of SET-MH monitoring stations showing elevation deficit, elevation surplus, and stability (parity) with contemporaneous RSLR. (**B**) Modeled probability of an elevation deficit with different rates of RSLR. (**C**) Histogram of paleo-marsh index points showing positive, negative, and no tendency in relation to RSLR in UK Holocene marshes (*13*). (**D**) Modeled probability of positive sea level tendency (i.e., sinking within the tidal frame) associated with different rates of Holocene RSLR. Numbers of observations for each RSLR increment are shown at the base of each column.

The US Atlantic and North Pacific coastlines have been subject to relatively high rates of RSLR for several millennia, forming organic-rich marshes situated within barrier and embayment geomorphic settings. Contemporary RSLR varies between coastlines as a result of vertical land movement and climate variability (*25*), with relatively high rates of RSLR occurring in the Gulf of Mexico, the US North Atlantic, and parts of the North Pacific coasts and lower RSLR in most European and Southern Hemisphere sites (data S1).

This variability in RSLR allowed us to identify RSLR influences on marsh surface elevation change, and on mechanisms driving the latter, including vertical accretion and shallow subsidence. SET-MH stations were monitored for an average of 10.1 years (range: 3.5 to 20.0 years) over periods for which RSLR at nearest tide gauges (hereafter "contemporaneous" RSLR: $\bar{x}$ = 6.81 ± 6.41 mm year^{-1}) was significantly higher than both the 50-year average ($\bar{x}$ = 3.75 ± 2.73 mm year^{-1}; $P < 0.001$) and the 3000-year average derived from glacio-isostatic modeling (*20*) ($\bar{x}$ = 0.65 ± 0.72 mm year^{-1}; $P < 0.001$). This range of RSLR values encompassed rates associated with marsh retreat in the Holocene stratigraphic record (*10*, *13*, *14*). We therefore tested three hypotheses concerning the feedback among RSLR, vertical accretion, and elevation gain: (i) The rate of vertical accretion would increase with RSLR; (ii) the rate of vertical accretion would correspond to sediment availability; and (iii) vertical accretion would correspond to marsh elevation gain. We determined the extent to which these relationships are influenced by climatic, environmental, and edaphic conditions (table S2), including soil bulk density and organic carbon.

The most important predictor of the rate of vertical accretion at a global scale was the 50-year RSLR trend (r^2 = 0.48, $P < 0.0001$). The observation of vertical accretion parity with increased RSLR aligns with the predictions of feedback models suggesting marsh resilience to RSLR (*7*, *8*, *12*), although the relationship was stronger in organic than in minerogenic marshes (fig. S1). Marsh accretion across the network was higher at sites that are lower in the tidal frame [Fig. 3B; as measured by dimensionless D (*20*), an indicator of submergence (*26*), $P < 0.0001$]. Annual average suspended matter in adjacent waters explained less than 10% of global-scale variability in vertical accretion (fig. S2), and the incorporation of tidal range as an additional variable [as has recently been suggested (*12*)] did not improve the prediction of the rate of vertical accretion (linear regression r^2 = 0.03; n = 410) or the r^2 [typically <0.01 for low marshes (i.e., $D > 0$; n = 168)], contrary to model projections (*5*, *7*, *12*). Vertical accretion on marsh surfaces in settings of low total suspended matter suggests an important role for accretion of autochthonous sediment (i.e.,

organic and/or locally resuspended mineral matter). One caveat is that satellite-derived measures of suspended matter may not represent sediment concentrations at the point of deposition, particularly in channelized estuarine settings. However, hydrogeomorphic setting was also not strongly predictive of the rate of vertical accretion (fig. S2).

Although vertical accretion was the most important control on surface elevation gain at the global scale (r^2 = 0.3; fig. S2), shallow subsidence mediates the relationship between vertical accretion and surface elevation gain (*27*, *28*) (Fig. 1 and Fig. 3C). Shallow subsidence was greater under higher accretion rates (r^2 = 0.34, $P < 0.0001$; Fig. 3C and fig. S3) and higher contemporaneous and 50-year RSLR (r^2 = 0.16, $P < 0.0001$; fig. S3). As a result, on average less than half of the sediment accreted above marker horizons translated into surface elevation gain, and this proportion decreased between 5 mm year^{-1} and 10 mm year^{-1} of contemporaneous RSLR ($P < 0.0001$). The deficit between surface elevation gain and RSLR trend increased linearly with RSLR in all settings (Fig. 3, E and F), as did the proportion of SET-MH monitoring stations subject to an elevation deficit (Fig. 4).

Marshes in the SET-MH network transition from a predominance of elevation surplus to elevation deficit over a similar range of RSLR, as has been historically observed in UK Holocene marshes (Fig. 4), which provide a record that is unique for the number of index points associated with a range of RSLR histories (*13*). Contemporary observations from the tidal marsh SET-MH network, which accounts for shallow subsidence, were therefore consistent with observations of tidal marsh and mangrove behavior during periods of relatively rapid sea level rise in the Holocene record (*10*, *13*, *14*, *22*). Cumulative probabilities based on Bayesian modeling using the SET-MH record suggest that a drowning trajectory is likely (66% probability) at 3.6 mm year^{-1} and 4.6 mm year^{-1} in UK Holocene marshes and very likely (90% probability) at 7.6 mm year^{-1} in the SET-MH record and 7.1 mm year^{-1} in UK Holocene marshes (*13*) (Fig. 4). Although several sites in the US Gulf and Atlantic coastlines had a contemporary rate of elevation gain exceeding 8 mm year^{-1}, these same sites had the lowest median projected time to open-water conversion, as estimated by the time to reach minimum survival elevation (*20*) (table S5). The elevation subsidy provided by their proximity to eroding shorelines (fig. S4) may represent laterally migrating levees (*29*), a precursor to marsh failure (table S4 and fig. S5).

In locations where sea level has been stable (<±0.5 mm year^{-1}) or falling over recent millennia (i.e., the macrotidal marshes of the North Atlantic and the Southern Hemisphere Australian and South African marshes), soil organic carbon concentrations were significantly lower (on average less than half) relative to marshes subject to millennial-scale RSLR ($P < 0.0001$; table S1 and fig. S6). Gradually rising sea levels can both promote and preserve highly organic marshes (*18*, *30*) by increasing plant productivity, increasing organic carbon burial, reducing oxidation, and slowing decomposition. At a global scale, the proportion of organic carbon in accreting sediments across our network was better explained by the 3000-year RSLR trend (r^2 = 0.23; $P < 0.0001$) than by contemporaneous RSLR (r^2 = 0.07; $P < 0.0001$; fig. S6). Sites with higher bulk density and lower percent organic carbon had lower rates of subsidence (table S1 and fig. S1) and a higher proportion of vertical accretion contributing to elevation gain, consistent with predictions (*9*). In these locations, shorelines were relatively stable and the proportion of vegetated to unvegetated marsh cover (*20*) was high (table S1).

The mechanisms promoting tidal marsh adjustment under low rates of sea level rise appear less effective under high rates of sea level rise. The substrates undergoing marsh elevation gain are increasingly subject to autocompaction and subsidence under increased accretion and inundation depth. The elevation response is nonlinear, and above a primary breakpoint between 5 and 10 mm year^{-1} of vertical accretion, a higher proportional loss to subsidence constrains elevation adjustment in response to accelerating RSLR. This observation reconciles the instrumental record with probabilities of tidal marsh adjustment emerging from paleostratigraphic records during phases of high RSLR during the Holocene. Both datasets suggest a low likelihood ($P < 0.1$) that tidal marshes will be retained in situ under global average rates of RSLR attained by mid-century under high-emissions scenarios, and by the end of the century under midrange-emissions scenarios (*25*). These rates of RSLR are already reached in subsiding deltas occupied by tidal marshes. Under these circumstances, tidal marsh survival will increasingly depend on their upland migration.

REFERENCES AND NOTES

1. N. C. Davidson, *Mar. Freshw. Res.* **65**, 934–941 (2014).
2. N. J. Murray *et al*., *Science* **376**, 744–749 (2022).
3. R. Costanza *et al*., *Glob. Environ. Change* **26**, 152–158 (2014).
4. S. Fagherazzi *et al*., *J. Geophys. Res.* **125**, e2019JF005200 (2020).
5. M. Schuerch *et al*., *Nature* **561**, 231–234 (2018).
6. M. L. Kirwan, J. P. Megonigal, *Nature* **504**, 53–60 (2013).
7. M. L. Kirwan, S. Temmerman, E. E. Skeehan, G. R. Guntenspergen, S. Fagherazzi, *Nat. Clim. Change* **6**, 253–260 (2016).
8. J. T. Morris, P. Sundareshwar, C. T. Nietch, B. Kjerfve, D. R. Cahoon, *Ecology* **83**, 2869–2877 (2002).
9. J. French, *Mar. Geol.* **235**, 119–136 (2006).
10. N. Saintilan *et al*., *Science* **368**, 1118–1121 (2020).
11. M. L. Kirwan *et al*., *Geophys. Res. Lett.* **37**, L23401 (2010).
12. D. J. Coleman *et al*., *Limnol. Oceanogr. Lett.* **7**, 140–149 (2022).
13. B. P. Horton *et al*., *Nat. Commun.* **9**, 2687 (2018).
14. T. E. Törnqvist, K. L. Jankowski, Y.-X. Li, J. L. González, *Sci. Adv.* **6**, eaaz5512 (2020).
15. M. L. Kirwan, G. R. Guntenspergen, *J. Geophys. Res.* **115**, F02009 (2010).
16. J. P. Syvitski, A. Kettner, *Philos. Trans. R. Soc. A* **369**, 957–975 (2011).
17. L. A. Deegan *et al*., *Nature* **490**, 388–392 (2012).
18. K. Rogers *et al*., *Nature* **567**, 91–95 (2019).
19. E. L. Webb *et al*., *Nat. Clim. Change* **3**, 458–465 (2013).
20. See supplementary materials.
21. C. E. Lovelock *et al*., *Nature* **526**, 559–563 (2015).
22. K. L. Jankowski, T. E. Törnqvist, A. M. Fernandes, *Nat. Commun.* **8**, 14792 (2017).
23. C. E. Lovelock, R. Reef, *One Earth* **3**, 195–211 (2020).
24. F. Wang *et al*., *Natl. Sci. Rev.* **8**, nwaa296 (2020).
25. B. Fox-Kemper *et al*., in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021), chap. 9.
26. J. T. Morris, *Estuar. Coast. Shelf Sci.* **69**, 395–402 (2006).
27. S. Nolte *et al*., *J. Coast. Conserv.* **17**, 301–325 (2013).
28. D. R. Cahoon, J. Day, D. Reed, *Curr. Top. Wetland Biogeochem.***3**, 72–88 (1999).
29. N. K. Ganju *et al*., *Nat. Commun.* **8**, 14156 (2017).
30. J. L. Breithaupt *et al*., *J. Geophys. Res.* **125**, e2019JG005349 (2020).

ACKNOWLEDGMENTS

A. Yeates contributed to data management. M. J. S. Casallas assisted with shoreline change detection. Data used in the paper are available as data S1. Original data for the Mississippi Delta and Mississippi Chenier sites were retrieved from the Coastal Information Management System (CIMS) database (http://cims.coastal.louisiana.gov). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US government. **Funding:** NRF/SAEON Elwandle Node; the NRF/DSI Shallow Marine and Coastal Research Infrastructure (SMCRI); the SAEON Elwandle Coastal Biogeochemical Platform: UKRI NERC CBESS (NE/J015423/1); Australian Research Council FS100100024, FT130100532, and the Coasts and Clean Seas Initiative; National Environmental Science Program 2 Coastal and Marine Hub; the DSI/NRF Research Chair in Shallow Water Ecosystems (UID 84375); US Geological Survey Climate Research and Development Program and Ecosystems Mission Area; Research Foundation of Flanders, Belgium (FWO grants G.0600.18N and G.0316.20N); NSF grants OCE-1702587 and OCE-2002437; Coastal Protection and Restoration Authority (CPRA) of Louisiana, 2021, Coastwide Reference Monitoring System-Wetlands Monitoring Data; US Geological Survey Western Ecological Research Center; Ministero delle Infrastrutture e dei Trasporti – Provveditorato Interregionale alle OO. PP. del Veneto – Trentino Alto Adige – Friuli Venezia Giulia, through its concessionary Consorzio Venezia Nuova; Catalan Water Agency, Ebro Delta Monitoring Network Program. **Author contributions:** Conceptualization, N.S., T.S., D.R.C., G.G.; data curation, N.S., K.R., N.C., G.G., J.C.L., D.R.C., J.A., J.R., K.E.K., T.S., D.A.F., T.M., P.M., S.T., C.E.L., K.W.K., G.L.C., J.B., C.I., F.S., K.T., J.A.G., E.A., K.C., C.E., C.M., P.M., K.S.L., W.R., K.B.R., T.B.; formal analysis, N.S., K.E.K., E.A., N.K., D.A.F., V.P.G.; project administration, N.S.; supervision, N.S., G.G.; writing–original draft, N.S.; writing–review and editing, N.S., K.E.K., G.G., K.R., J.C.L., D.R.C., C.E.L., D.A.F., E.A., K.W.K., N.C., T.S., J.A., J.R., C.I., F.S., S.T., P.M., T.M., K.T., J.B., G.L.C., T.B., V.P.G., K.C., C.E., C.M., K.S.L., W.R., K.B.R., J.A.G. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Original SET-MH data held by the USGS are available at doi:10.5066/P9UZ4LQ2.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo7872
Materials and Methods
Figs. S1 to S6
Tables S1 to S5
References (*31*–*65*)
Data S1

Submitted 24 February 2022; accepted 27 June 2022
10.1126/science.abo7872

SPECIES INTERACTIONS

Pollinators of the sea: A discovery of animal-mediated fertilization in seaweed

E. Lavaut[1], M.-L. Guillemin[1,2], S. Colin[3,4], A. Faure[1], J. Coudret[1], C. Destombe[1], M. Valero[1]*

The long-held belief that animal-mediated pollination is absent in the sea has recently been contradicted in seagrasses, motivating investigations of other marine phyla. This is particularly relevant in red algae, in which female gametes are not liberated and male gametes are not flagellated. Using experiments with the isopod *Idotea balthica* and the red alga *Gracilaria gracilis*, we demonstrate that biotic interactions dramatically increase the fertilization success of the alga through animal transport of spermatia on their body. This discovery suggests that animal-mediated fertilization could have evolved independently in terrestrial and marine environments and raises the possibility of its emergence in the sea before plants moved ashore.

Animal-mediated pollination in plants is believed to have originated more than 140 million years ago (*1*) [although this date still sparks debate (*2*)]. Insect pollination is the most common form of fertilization in flowering plants (*3*) and is usually the unintended consequence of an animal's feeding activity. Although extensively studied on land, a demonstration of animal pollination in the marine environment was made in 2012 when foraging marine invertebrates were shown to carry and transfer pollen grains from male to female flowers in the seagrass *Thalassia testudinum* K.D. Koenig (*4*, *5*). This finding contradicted the long-held belief that animal-mediated pollination is absent in the sea (*6*) and raised the question of its presence in other marine life, particularly in seaweeds.

We investigated whether animal-mediated dissemination of male gametes occurs in seaweed. This question is particularly relevant in Florideophyceae because male gametes are not flagellated or motile, similar to pollen grains, and fertilization takes place on a remote female organ. The sperm (spermatia) cannot swim to attach to the trichogyne (the female receptive structure) and are typically thought to be dispersed only passively through water flows (*7*).

The marine isopod *Idotea balthica* (Pallas, 1772) is commonly found firmly gripped to the red alga (Fig. 1A and movie S1) *Gracilaria gracilis* (Stackhouse) M. Steentoft, L.M. Irvine & W.F. Farnham. Using two sets of experiments, we tested whether *I. balthica* facilitates male gamete dispersal and fertilization in this red alga species either indirectly, through the creation of water turbulence when swimming, or directly, by carrying spermatia on their bodies. Mating in *G. gracilis* occurs when spermatia released into the environment (Fig. 1B) encounter a trichogyne on the female thallus, where fertilization takes place. After fertilization, the zygote develops on the female alga into a complex structure visible to the naked eye called a cystocarp (fig. S1). We therefore used the number of cystocarps per centimeter of thallus as a proxy of fertilization success.

In the first set of experiments, virgin reproductive-female thalli were placed in calm seawater aquaria 15 cm apart from reproductive-male thalli, either in the presence or absence of idotea isopods (Fig. 2, A and B, respectively). Fertilization success was about 20 times higher in the presence of idoteas than without (Wilcoxon rank-test, $n = 5$ replicates, $P = 0.031$) (Fig. 2C). In the second set of experiments, virgin females were placed in aquaria with or without idoteas that had been preincubated with reproductive-male thalli (Fig. 2, D and E, respectively). To control for the possibility that sperm contamination by the operator during animal transfer led to fertilization events, we performed at the beginning of the experiment shown in Fig. 2E a mock transfer that involved plunging the instruments used to handle idoteas. Only two cystocarps were detected—and in only one replicate—after the mock transfer (an average 0.002 cystocarp/cm) (table S1), indicating that such sperm contamination by the operator was negligible. Fertilization success rate reached 0.11 cystocarp/cm of thallus in the presence of incubated idoteas (Wilcoxon rank-test, $n = 5$ replicates, $P = 0.028$) (Fig. 2F), confirming direct gamete transport by isopods. Idoteas were imaged by use of confocal laser scanning microscopy to demonstrate that they were indeed carrying spermatia. The presence of spermatia was revealed with staining (Fig. 1C). The cuticles of the animals were studded with spermatia (Fig. 1D, fig. S2, and movie S2), mainly adhering to the setae of pereiopods (Fig. 1E and movie S2) and on the margin segments (Fig. 1F and movie S2) because of the sticky nature of the sperm mucilage (fig. S3).

Our results demonstrate for the first time that biotic interactions dramatically increase the probability of fertilization in a seaweed. The observation of mucilage-embedded spermatia attached to idotea body parts after visiting male gametophytes suggests that idotea may serve as "pollinators" in *G. gracilis* upon subsequent visits to female individuals (movie S1). If we draw a parallel with the Cox-Knox postulates that define criteria for determination of pollinators of flowering plants (*8*), our results are sufficient to prove that idoteas are efficient gamete vectors.

This discovery is particularly relevant in Rhodophyta because male gametes are not flagellated and have a short life span (*9*), and more than 75% of species are dioicous (having male and female structures on different individuals), which means that fertility success is distance dependent and limited by mate density (*10*). Syngamy in red algae is expected to be relatively rare compared with that of marine organisms that possess external fertilization, the synchronous release of motile gametes, and chemotaxis (such as invertebrates and brown seaweeds), unless compensatory mechanisms increase sperm production, sperm transfer, or zygote survival. Searles hypothesized that the evolution of a life history in which the female nurtures and protects the zygote is probably a response to low rates of fertilization in both plants and red algae (the cystocarps) (*7*). Despite debate concerning the compensatory allocation of resources for zygote survival in red algae, the possibility of a biotic interaction that facilitates fertilization success in this group had not been previously explored. In particular, animal-mediated transport of spermatia in red algae could dramatically increase fertilization success in calm rock pools (a condition that was mimicked by our experimental conditions) at low tide, when most fertilizations have been detected (*11*). However, the relative importance of animal-mediated transport versus water movements in fertilization success was not tested. Intertidal seaweed could combine the advantages of both water and animal dispersal of male gametes when alternating between low- and high-tide conditions.

Although *I. balthica* grazes on other seaweed, it does not feed directly on *Gracilaria* but rather eliminates epiphytes at the surface of the thallus while protecting itself from predators (*12*–*14*). We suggest that the relationship could be mutually beneficial. For *I. balthica*, the seaweed provides shelter, and epiphytic diatoms found adhering to the surface of the thallus, whose frustules (or thecae)

[1]International Research Laboratory (IRL) 3614, Evolutionary Biology and Ecology of Algae, Centre National de la Recherche Scientifique (CNRS), Sorbonne Université, Pontificia Universidad Católica de Chile, Universidad Austral de Chile, CS 90074, Place Georges Teissier, 29688 Roscoff, France. [2]Instituto de Ciencias Ambientales y Evolutivas, Núcleo Milenio MASH, Facultad de Ciencias, Universidad Austral de Chile, 567 Valdivia, Chile. [3]Unité Mixte de Recherche (UMR) 8227, Laboratory of Integrative Biology of Marine Models, CNRS, Sorbonne Université, CS 90074, Place Georges Teissier, 29688 Roscoff, France. [4]BioOptics facility, Max Planck Institute (MPI) for Biology Tübingen, Max-Planck-Ring 5, 72076 Tübingen, Germany.
*Corresponding author. Email: myriam.valero@sb-roscoff.fr

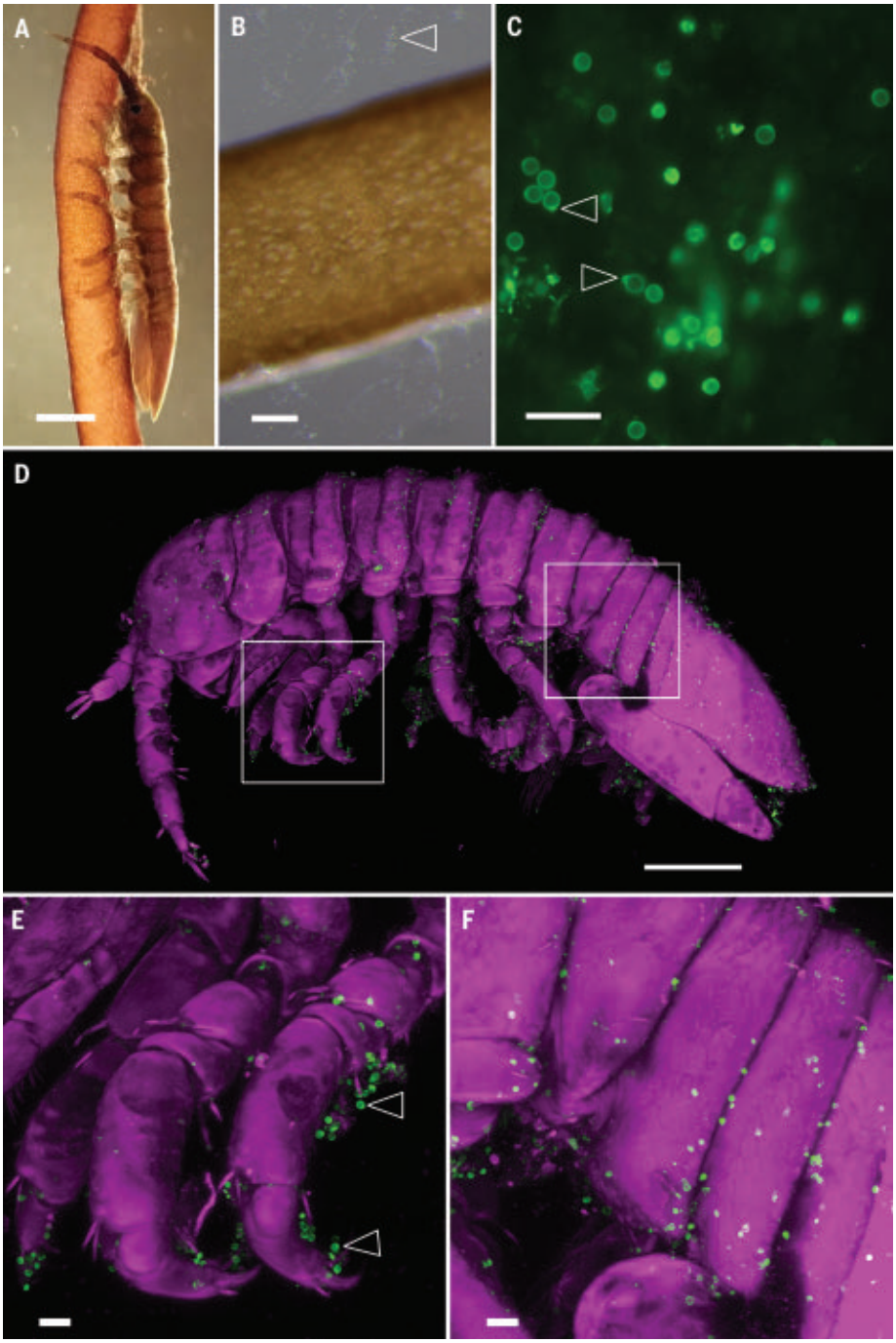


Fig. 1. The idoteas carry spermatia of *G. gracilis* on their bodies. (**A**) *I. balthica* firmly clinging to the reproductive-male thallus of *G. gracilis* (oblique illumination, macroscopy) (movie S1). The light spots on the thallus correspond to the male conceptacles. (**B**) Release of spermatia (arrowhead) at the surface of the reproductive-male thallus (oblique illumination, macroscopy). (**C**) Spherical spermatia about 4.5 μm in diameter (arrowheads) with their noticeable bump [epifluorescence microscopy, staining with the lectine Concanavalin A conjugated to the fluorochrome (AlexaFluor488); ConA-AF488 binds to specific carbohydrates (*24*) on the surface of the cell and mucilage]. (**D**) Three-dimensional reconstruction from confocal laser scanning microscopy of a young specimen of *I. balthica* collected on a reproductive-male gametophyte of *G. gracilis* (tegument of the isopod was fixed and stained with calcofluor white, and the spermatia was stained with ConA-AF488 (figs. S2 and S3 and movie S2). (**E**) Close-up of (D) right inset, showing location of spermatia on the setae of pereiopods. (**F**) Close-up of (D) left inset, showing location of spermatia on the margin of segments (arrows). Scale bars, (A) 1 mm; (B) and (D) 200 μm; and (C), (E), and (F) 20 μm.

are found in the feces of idoteas (fig. S4), appear to be an important food source. In return, *G. gracilis* benefits through increased growth rate owing to reduced fouling (*15*) and improved reproductive success. The benefit of animal-mediated "pollination" may be particularly important in species such as *G. gracilis* with funnel-shaped trichogynes that do not protrude from the surface of the thallus (*16*), making fertilization more difficult compared with that of other Rhodophyta such as *Bostrychia moritziana* (*17*), which have elongated trichogyne (about 100 μm long). As in flowering plants in which ambophily (adaptation to both wind and animal pollination) appears to be more important than previously reported (*18*), our results demonstrate that in addition to the well-established role of water movement (*7*, *9*, *10*), animal-mediated fertilization is also present in the sea. This observation opens up new unexplored avenues of research aimed at understanding the ecology of these interactions.

Animal pollination is generally thought to have originated during the Mesozoic for gymnosperms and angiosperms and was believed absent in mosses and ferns, in which fertilization is restricted to water (*19*–*21*). However, transport of male gametes by animals has been demonstrated recently in mosses, antedating animal-mediated pollination at the early phase of land colonization (~ 450 million years ago) (*21*). The discovery of animal-mediated fertilization in Florideophyceae, whose origin was estimated to be between 817 million and 1049 million years ago (*22*), suggests that it could have arisen with the diversification of the metazoan some 650 million years ago (*23*), even before plants moved ashore. However, we cannot rule out that different animal-mediated fertilization

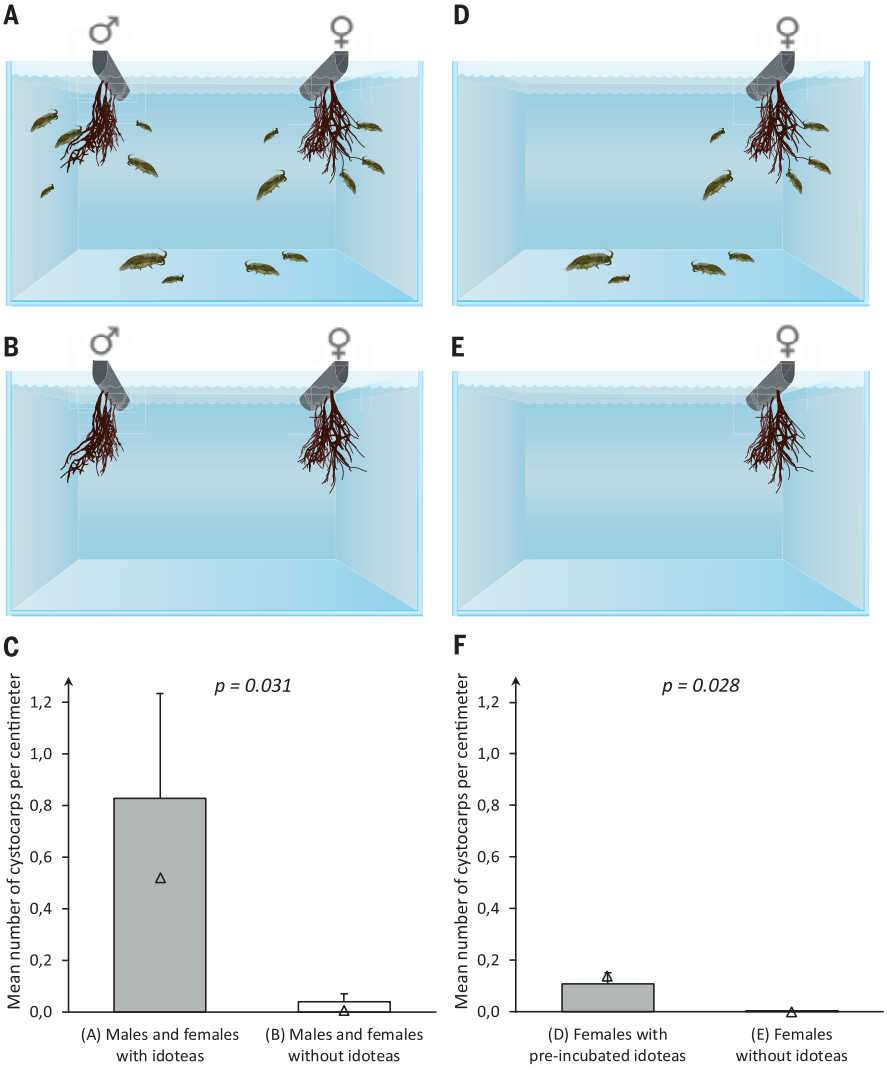


Fig. 2. Impact of idoteas on fertilization success. (**A** and **B**) Aquarium containing (A) male and female *Gracilaria gracilis* with 20 *Idotea balthica* (movie S1) and (B) male and female *G. gracilis* without *I. balthica.* (**C**) Comparison of the mean number of cystocarps as a proxy of fertilization success between experiments shown in (A) and (B). (**D**) Females of *G. gracilis* with 20 *I. balthica* preincubated with male gametophytes. (**E**) Negative control with females of *G. gracilis* and mock transfer of idoteas. (**F**) Comparison of the mean number of cystocarps between experiments (D) and (E). Values represented on the graph are the means and the standard error for the five replicates. Medians are indicated with a triangle. *P* values are indicated on the graph (unilateral rank test of Wilcoxon for paired samples). [Aquarium illustrations by E. Mardones]

mechanisms evolved independently and repeatedly in terrestrial and marine environments.

REFERENCES AND NOTES

1. D. A. Grimaldi, M. S. Engel, *Evolution of the Insects* (Cambridge Univ. Press, 2005).
2. C. J. van der Kooi, J. Ollerton, *Science* **368**, 1306–1308 (2020).
3. J. D. Ackerman, *Plant Syst. Evol.* **222**, 167–185 (2000).
4. B. I. van Tussenbroek, L. Monroy-Velazquez, V. Solis-Weiss, *Mar. Ecol. Prog. Ser.* **469**, 1–6 (2012).
5. B. I. van Tussenbroek *et al*., *Nat. Commun.* **7**, 12980 (2016).
6. G. J. Vermeij, R. K. Grosberg, *Integr. Comp. Biol.* **50**, 675–682 (2010).
7. R. B. Searles, *Am. Nat.* **115**, 113–120 (1980).
8. J. Ollerton, *Pollinators and Pollination: Nature and Society* (Pelagic Publishing, 2021).
9. C. Destombe, J. Godin, J.-M. Remy, *Hydrobiologia* **204–205**, 219–223 (1990).
10. C. R. Engel, R. Wattier, C. Destombe, M. Valero, *Proc. Biol. Sci.* **266**, 1879–1886 (1999).
11. C. R. Engel, C. Destombe, *J. Mar. Biol. Assoc. U. K.* **82**, 189–192 (2002).
12. T. M. Bell, E. E. Sotka, *Oecologia* **170**, 383–393 (2012).
13. C. F. Aumack, C. D. Amsler, J. B. McClintock, B. J. Baker, *J. Phycol.* **47**, 36–41 (2011).
14. C. D. Amsler, J. B. McClintock, B. J. Baker, *J. Phycol.* **50**, 1–10 (2014).
15. G. Mancinelli, L. Rossi, *Mar. Biol.* **138**, 1163–1173 (2001).
16. S. Fredericq, M. H. Hommersand, *J. Phycol.* **25**, 213–227 (1989).
17. S. M. Wilson, J. D. Pickett-Heaps, J. A. West, *Eur. J. Phycol.* **37**, 509–522 (2002).
18. C. Larue, E. Austruy, G. Basset, R. J. Petit, *Bot. Lett.* **168**, 348–372 (2021).
19. S. D. Klavins, D. W. Kellogg, M. Krings, E. L. Taylor, T. N. Taylor, *Evol. Ecol. Res.* **7**, 479–488 (2005).
20. T. Bao, B. Wang, J. Li, D. Dilcher, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24707–24711 (2019).
21. N. Cronberg, R. Natcheva, K. Hedlund, *Science* **313**, 1255–1255 (2006).
22. E. C. Yang *et al*., *Sci. Rep.* **6**, 21361 (2016).
23. D. H. Erwin *et al*., *Science* **334**, 1091–1097 (2011).
24. S. H. Kim, G. H. Kim, *Hydrobiologia* **398/399**, 81–89 (1999).

ACKNOWLEDGMENTS

We thank T. Broquet, T. Comtet, P. H. Gouyon, M. Hossaert, C. Houbin, D. Roze, and R. D. Sepúlveda for fruitful discussion on the first steps of this study; S. Loisel for his help in construction of the aquaria; and S. Mauger for sequencing. We thank the light microscopy facilities from the Station Biologique de Roscoff (SBR) and the MPI for Biology Tübingen for providing access to confocal laser scanning microscopes. We thank G. Schires and the Center of Marine Biology Resources facilities from the SBR for providing access to the aquariological structure for seaweed cultivation. We thank D. Inouye, J. Ollerton, and an anonymous reviewer for comments that have improved this article. **Funding:** This work was supported by a Centre National de la Recherche Scientifique (CNRS) PhD fellowship to E.L.; core funding to M.V., C.D., and M.-L.G. from the IRL 3614, Evolutionary Biology and Ecology of Algae, CNRS and Sorbonne Université; the International Research Network (IRN), Diversity and Biotechnology of Marine Algae, CNRS; and Gobierno de Chile, Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) #1170541 to M.-L.G. **Author contributions:** E.L., C.D., M.-L.G., and M.V. designed the experiment; E.L., C.D., A.F., and M.V. performed the field sampling; E.L., J.C., and A.F. conducted the laboratory work; E.L., A.F., and M.V. performed the statistical analyses; S.C. designed and conducted the CLSM experiment, analyzed the imaging data, and prepared the corresponding figures; E.L., S.C., C.D., M.-L.G., and M.V. wrote and revised the manuscript, with further edits from all authors; M.-L.G. and M.V. obtained the funding; and M.V. designed and supervised the project, with C.D. and M.-L.G. **Competing interests:** The authors declare no competing interests. **Data availability:** All data used in the analyses are available in the paper or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0661
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References (*25–30*)
MDAR Reproducibility Checklist
Movies S1 to S2

Submitted 17 February 2022; accepted 16 June 2022
10.1126/science.abo6661

SOLAR CELLS

Inactive $(PbI_2)_2RbCl$ stabilizes perovskite films for efficient solar cells

Yang Zhao[1,2], Fei Ma[1,2], Zihan Qu[1,2], Shiqi Yu[1,2], Tao Shen[2,3], Hui-Xiong Deng[2,3], Xinbo Chu[1,2], Xinxin Peng[4], Yongbo Yuan[4,5], Xingwang Zhang[1,2], Jingbi You[1,2]*

In halide perovskite solar cells the formation of secondary-phase excess lead iodide (PbI_2) has some positive effects on power conversion efficiency (PCE) but can be detrimental to device stability and lead to large hysteresis effects in voltage sweeps. We converted PbI_2 into an inactive $(PbI_2)_2RbCl$ compound by RbCl doping, which effectively stabilizes the perovskite phase. We obtained a certified PCE of 25.6% for $FAPbI_3$ (FA, formamidinium) perovskite solar cells on the basis of this strategy. Devices retained 96% of their original PCE values after 1000 hours of shelf storage and 80% after 500 hours of thermal stability testing at 85°C.

Several strategies can enhance the power conversion efficiency (PCE) of perovskite solar cells (PSCs), such as growing films with high crystallinity (*1*, *2*), doping with anions, cations, or both (*3*–*5*), engineering the charge transport layer (*6*–*12*), and use of a universal passivation layer (*13*–*16*). Several groups have reported certified PCEs >25% by modulating the SnO_2 electron transport layer (*10*–*12*).

Previous reports have shown that the secondary phase of PbI_2 is critical for attaining high performance levels (*7*, *12*, *17*–*24*). The benefits of excess PbI_2—either in the bulk or at the grain boundary of perovskite films—have mainly been attributed to its passivation effect resulting from the formation of a type-I band alignment (*24*). However, excess PbI_2 is too active and leads to instability of the device and large hysteresis in current density voltage (*J-V*) characteristics (*18*, *23*, *25*, *26*). Both factors can be ascribed to the photodecomposition of PbI_2 and the enhanced ionic migration.

It is crucial to manage excess unstable PbI_2 in perovskite films to achieve high efficiency. Recently, Luo *et al.* developed a ligand-modulation approach to manage the excess PbI_2 and successfully improved both PCE (up to 22%) and stability (*23*).

We propose stabilization of the perovskite by converting excess PbI_2 into an inactive, new, secondary-phase $(PbI_2)_2RbCl$ that can improve the stability of the perovskite $FAPbI_3$ layer (for simplicity, the perovskite film with $(PbI_2)_2RbCl$ is henceforth referred to as PIRC), in addition to improving its optoelectronic properties, the reduction of PbI_2 decreased the bandgap of the perovskite layer. As a result, a 26.1% efficiency (certification efficiency: 25.6%) of perovskite solar cells with >1000 hours of shelf stability and 500 hours of thermal stability at 85°C have been developed.

We deposited nearly pure $FAPbI_3$ by a modified two step method used in our previous reports (*7*, *15*, *18*). For deposition of $FAPbI_3$, the PbI_2 seed layer was first grown on the substrate while the organic salt formamidinium iodide (FAI) solution was deposited on PbI_2 for diffusion and formation of perovskite by annealing. Further, to improve the perovskite crystallization, methylammonium chloride (MACl) was added to the FAI solution (*7*, *27*). In contrast to our previous reports, to enhance the thermal stability and decrease the bandgap of perovskite as much as possible, traditional organic salt methylammonium iodide (MAI) or methylammonium bromide was not used in this study. We confirmed that a 5% molar ratio of RbCl doping is the best condition for device performance, as discussed below.

We performed scanning electron microscopy (SEM) (Fig. 1, A and B) measurements to determine the perovskite film crystallinity. For the control film, the perovskite crystal grain size was ~1 μm and the grains were surrounded by PbI_2 grains (white needle shapes, Fig. 1A), which formed as decomposition products on the perovskite surface after heating at elevated temperatures. For the PIRC films, the perovskite crystal size was increased to ~2 μm. Instead of a large fraction of evenly distributed white needle-shaped grains, sporadically distributed white irregular flakes were observed (Fig. 1B). Energy dispersive spectroscopy mapping of the PIRC film (fig. S1) revealed that the white region contained Pb, I, Rb, and Cl, but no N. This result indicated that there was

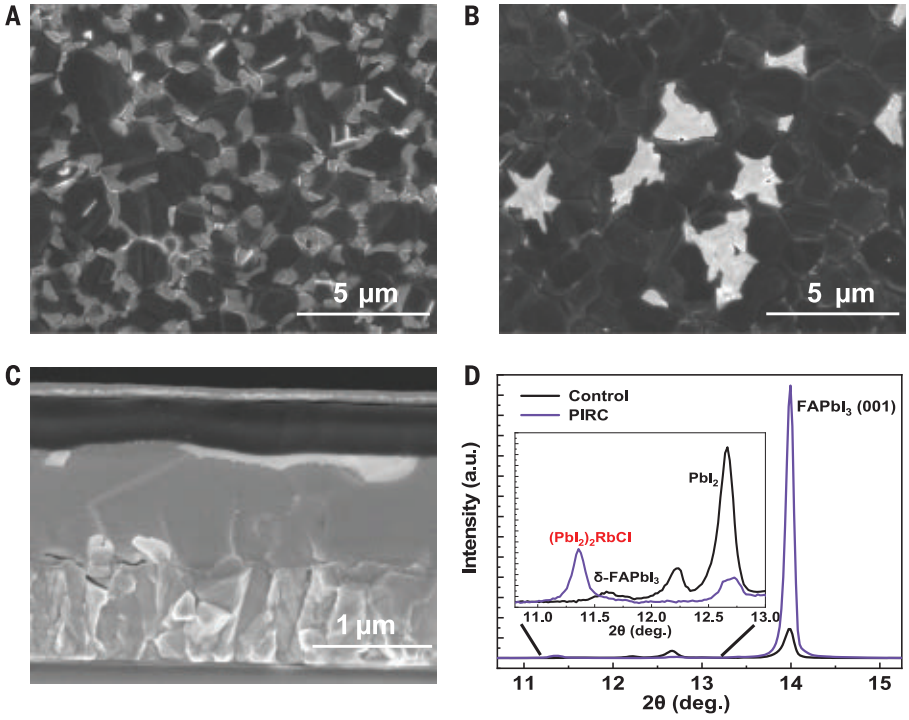


Fig. 1. Microstructures and morphology of secondary-phase $(PbI_2)_2RbCl$ on perovskite. (**A** and **B**) SEM of perovskite films without RbCl and with 5% RbCl, respectively. Scale bars, 5 μm. (**C**) Cross-section SEM image of a completed device using RbCl doped perovskite. Scale bar, 1 μm. (**D**) XRD of perovskite without RbCl and with 5% RbCl.

[1]Key Laboratory of Semiconductor Materials Science, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, P.R. China, 100083. [2]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing, P.R. China, 100049. [3]State Key Laboratory of Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, P.R. China, 100083. [4]Hunan Key Laboratory of Super Microstructure and Ultrafast Process, School of Physics and Electronics, Central South University, Changsha, Hunan, P.R. China, 410083. [5]Hunan Key Laboratory of Nanophotonics and Devices, School of Physics and Electronics, Central South University, Changsha, Hunan, P.R. China, 410083.
*Corresponding author. Email: jyou@semi.ac.cn

no organic-inorganic hybrid $FAPbI_3$ perovskite phase in this region whereas a PbI_2/RbCl mixed phase was possible (fig. S1 and tables S1 and S2). The cross-section SEM images showed that the new secondary phases only formed on top of the perovskite films (Fig. 1C).

Conventional x-ray diffraction (XRD) for the perovskite films (Fig. 1D) showed that the main diffraction peak of the control film could be assigned from the black phase of $FAPbI_3$ located at 14°. The yellow phase of $FAPbI_3$ (11.7°) was also observed in addition to the PbI_2 diffraction peak (12.6°) (the inset of Fig. 1D shows the enlarged figure). The PbI_2 diffraction peak in the PIRC film was largely suppressed and a diffraction peak (11.3°) emerged (Fig. 1D, inset). Grazing-incidence wide-angle x-ray scattering (GIWAXS) of the PIRC films were also consistent with conventional XRD results (fig. S2). We directly compared the XRD pattern of the annealed PbI_2/RbCl precursor films with different ratios (fig. S3), and we found that the diffraction of PbI_2 almost disappeared whereas the diffraction intensity at 11.3° was greatly enhanced when the ratio of PbI_2/RbCl is 2. Thus, we could attribute the 11.3° peak to crystalline $(PbI_2)_2RbCl$. Density function theory calculations and powder diffraction results also confirmed this assignment (fig. S4).

To study the phase stability of the perovskite materials obtained we aged the perovskite films at 85°C for 48 hours. The XRD and SEM results showed that more PbI_2 was formed after heating of the control perovskite film (figs. S5 and S6). We found that integration of the XRD intensity ratio of PbI_2/perovskite in the control film increased from 0.42 to 0.98 after aging whereas the PIRC film remained at ~0.10. These results indicate that excess PbI_2 can

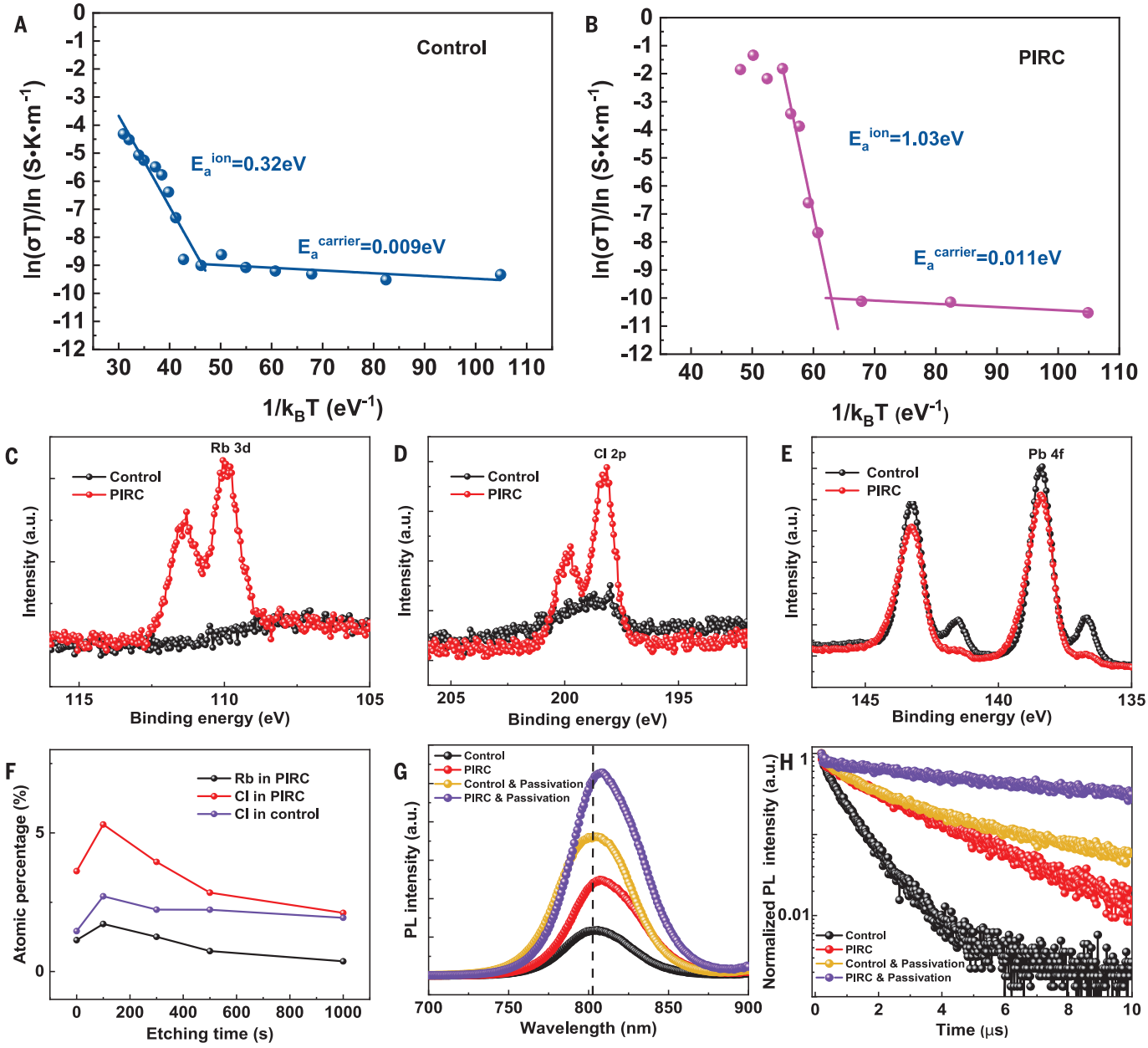


Fig. 2. Properties of control and PIRC perovskite films including ion transport, composition, and photoluminescence. (**A** and **B**) Temperature-dependent conductivity measurements of control and PIRC perovskite films, respectively. (**C** to **E**) XPS spectra of Rb 3d, Cl 2p, and Pb 4f core energy levels in control and PIRC films, respectively. (**F**) XPS depth profile of Cl in the control and PIRC samples; the Rb depth in the PIRC sample is also included. (**G** and **H**) Steady PL and TRPL of control perovskite film, PIRC film, and their surface passivation films.

trigger decomposition of the perovskite layer (*18*, *25*, *26*, *28*, *29*) whereas the perovskite becomes much more stable after conversion of PbI_2 into the $(PbI_2)_2RbCl$ phase.

The presence of PbI_2 in the perovskite layer could lead to ion migration as a result of the formation of Schottky defects such as FA and I vacancies (*30*). We measured the temperature-dependent conductivity with a previously reported procedure (*31*) to confirm that ion migration is suppressed in the PIRC films through reduction of PbI_2 (Fig. 2, A and B). We can estimate the ion migration active energy based on the Nernst-Einstein relation:

$$\sigma T = \sigma_0 \exp(-E_a/k_B T)$$

where σ_0 is a constant, k_B is Boltzmann's constant, σ is ionic conductivity, T is temperature, and E_a is the ion-migration activation energy. We derived E_a values from the slope of the ln (σT) versus $1/T$ of ~0.32 eV for the control films and ~1.03 eV for the PIRC films. Given the E_a, we can estimate that the ion migration rates (k) at the operating temperature (50°C) are 4.2×10^8 s^{-1} and 3.3×10^{-3} s^{-1}, respectively, based on the Arrhenius equation:

$$k = (k_B T)/h \exp(-E_a/RT)$$

where k_B stands for the Boltzmann constant, T is temperature, h is the reduced Planck constant, and R is the ideal gas constant (*32*). We found that $(PbI_2)_2RbCl$ could not easily react with FA cations or I ions compared with those of PbI_2 (fig. S7), inferring that $(PbI_2)_2RbCl$ lacks similarity to PbI_2 in regards to acting as a sink for absorbing adjacent FA and I during heating, photo soaking, or voltage application. This will suppress FA and I vacancy formation, as well as ion migration. Another possible reason for suppression of ion migration as a result of $(PbI_2)_2RbCl$ is that as the derivative of PbI_2, the $(PbI_2)_2RbCl$ is a fully intercalated compound (fig. S4) with narrow ion migration pathways whereas PbI_2 is a layer structure crystal with enough interlamellar space for ion migration.

X-ray photoelectron spectroscopy (XPS) characterization of the PIRC film showed the expected Rb 3d core level peak (Fig. 2C). For Cl, a substantial difference was found between the control and PIRC films. We estimated ~2.7% Cl in the control film (Fig. 2D), resulting from the addition of residual MACl during perovskite film growth, although most MACl would have evaporated during annealing. In these perovskite films, the radius of Cl was too small to incorporate into the lattice. We estimated that ~5.3% Cl was left in the PIRC film that was bonded with Pb or Rb ions and could not be easily evaporated (Fig. 2D). This excess Cl could stabilize the perovskite phase through strong Pb-Cl bonding (*8*–*10*). The presence of RbCl also enhanced the formation of the black

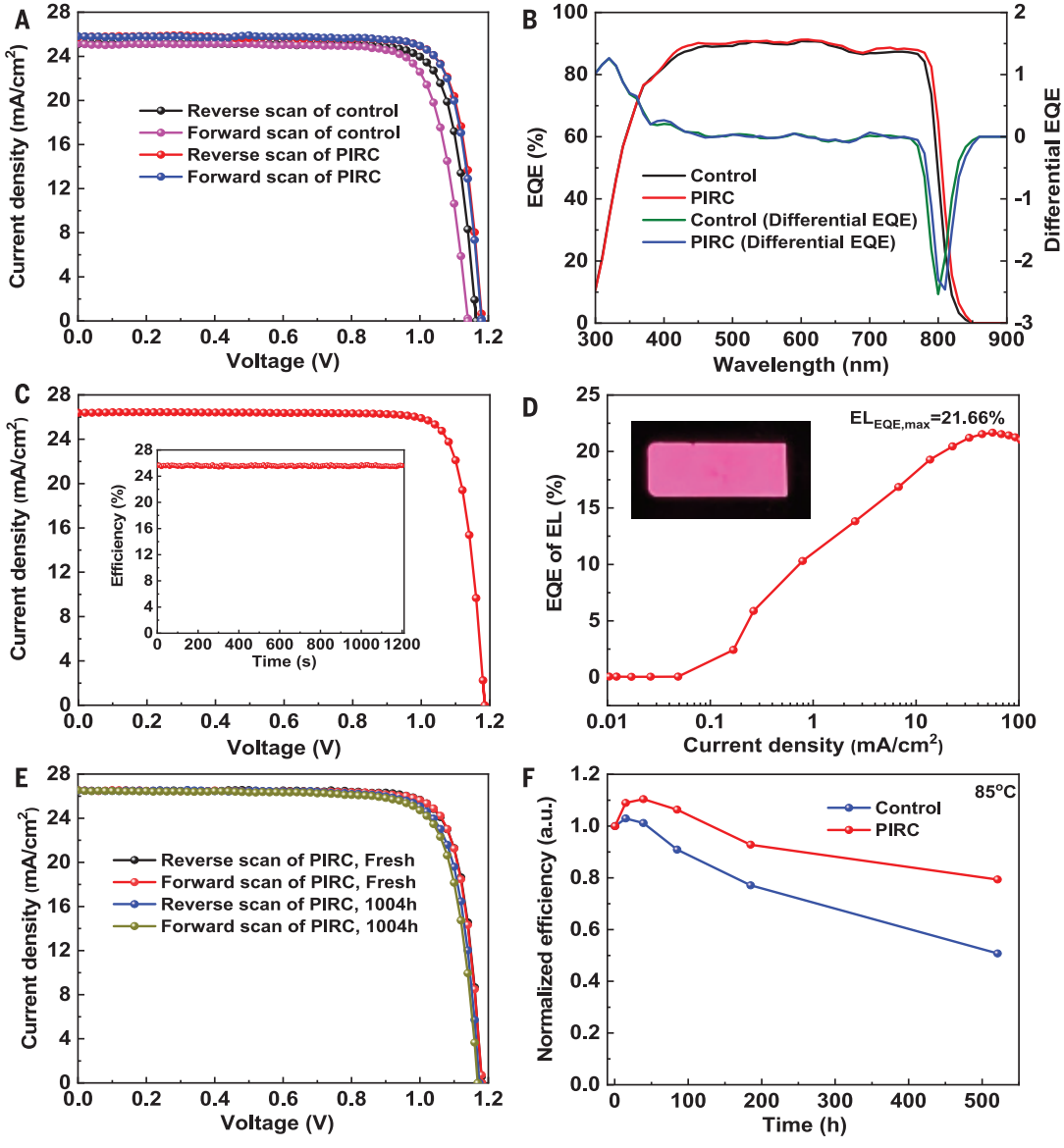


Fig. 3. Device characteristics and accelerated stability tests. (**A**) The typical *J-V* curves of the control and PIRC-based devices under one-sun (100 mW/cm^2) conditions. Both the reverse and forward scans are included. (**B**) EQE of devices for control and PIRC-based devices. (**C**) Best device performance based on PIRC films. (Inset) MPP of the device. (**D**) EQE of electroluminescence of the PIRC-based devices while operating as light-emitting diodes. (Inset) EL image of the device. (**E**) *J-V* curves of the PIRC-based device before and after aging in a nitrogen environment for 1004 hours. (**F**) Thermal stability (85°C) of the control and PIRC-based devices.

perovskite phase (Fig. 1D). Previous studies showed that the metallic Pb formed from PbI_2 under high-energy illumination drives loss of I_2 (*28*, *29*). We found that the metallic state of Pb was suppressed after introducing PIRC (Fig. 2E), which could be due to much better photo stability of PIRC compared with PbI_2 (fig. S8).

Depth profile XPS for the PIRC films revealed that the top surface was Rb- and Cl-rich (Fig. 2F and fig. S9), which could be a result of the secondary phase having a different solubility in the precursor solvent. Time-of-flight second ion mass spectroscopy (ToF-SIMS) of the control and target perovskite films (fig. S10) also confirmed that Rb and Cl that convert PbI_2 into $(PbI_2)_2RbCl$ were mainly distributed on top of the perovskite. The Cl ratio on the bottom of the perovskite is about 1% higher than that in the bulk estimated from ToF-SIMS (fig. S10) and XPS results (Fig. 2F).

Excess PbI_2 can increase the perovskite bandgap (*7*, *12*, *17*), which could be attributed to the strong quantum confinement in perovskite/PbI_2. The photoluminescence (PL) emission of the control film peak at 805 nm red shifted to 810 nm in the PIRC film (Fig. 2G). The PIRC perovskite film showed enhanced PL intensity indicative of suppression of nonradiative recombination. Time-resolved photoluminescence (TRPL) showed that the lifetime of the perovskite increased from 0.98 μs to 2.3 μs for the PIRC films and that this lifetime can be extended to ~6 μs with the addition of a passivation layer on the perovskite surface (Fig. 2H).

We fabricated the solar cell devices in an n-i-p configuration of FTO/SnO_2/perovskite/passivation layer/Spiro-OMeTAD/Au [FTO, fluorine-doped tin oxide; spiro-OMeTAD, 2,2′,7,7′-tetrakis(N,N-dipmethoxyphenylamine)-9,9′-spirobifluorene; Fig. 1C]. After testing the device performance using perovskite films with 1 to 10% RbCl doping we found that the 5% RbCl device had the best performance (fig. S11 and table S3); additionally, improved performance was observed when Cl was used as the halide (fig. S12 and table S4). A representative *J-V* curve for the control and 5% RbCl-doped device showed that each of the photovoltaic parameters increased (Fig. 3A). The typical short circuit current density (J_{SC}) increased from 25.1 to 25.8 mA/cm^2, which could be attributed mainly to the lower bandgap (Fig. 2G) as well as the external quantum efficiency (EQE) (Fig. 3B). The open circuit voltage V_{OC} increase from 1.17 to 1.19 eV despite the bandgap decrease was due to the suppression of recombination. The fill factor (FF) was almost the same or slightly improved compared with the control device.

Overall, the PCE increased from 24.6 to 25.6%. In addition, the PIRC-based device was almost hysteresis-free whereas the control device showed obvious hysteresis. The highest PCE for an optimized device tested in our laboratory was 26.1% (Fig. 3C); said device also showed 25.7% efficiency and steady output for 1200 s (device area: 0.108 cm^2; measurement mask size: 0.0741 cm^2). The statistics of the PCE distribution confirm the reproducibility of our results (fig. S13). These devices, which showed 25.6% certificated efficiency with a V_{OC} of 1.182 V, a J_{SC} of 26.3 mA/cm^2, and a *FF* of 82.7% (figs. S14 and S15), were sent to the accredited laboratory in Newport, USA, for certification.

The V_{OC} reflects the recombination rate of a given device and can be measured through the EQE at short circuit current conditions when the device is tested as a light-emitting diode. These devices had electroluminescence (EL) efficiencies as high as 21.6% under the injection current of 26 mA/cm^2 (equal to short circuit current J_{ph}) (Fig. 3D). High quantum efficiency can only be observed through charge confinement in light-emitting diodes, so the high-EL efficiency indicates that the defect-related nonradiative recombination was largely suppressed in the tested device. Light intensity–dependent results (fig. S16) also confirm this conclusion (*33*) and are consistent with our device simultaneously delivering high J_{sc} and V_{oc}.

We tracked device shelf stability and hysteresis in a nitrogen environment while using the Spiro-OMeTAD as the hole transport layer. After 1004 hours of aging, a slight drop in the FF from 82.6 to 80.6% was observed for the RPIC strategy–based device, which could be attributed to the reduction of Spiro-OMeTAD in an oxygen-free environment (*34*, *35*); further, it had 96% of its original PCE. The device maintained almost the same PCE in both reverse and forward scans (Fig. 3E). However, the control device could only maintain its reverse scan PCE of 24.1% whereas the forward efficiency was reduced to 22.9%. We attributed this enhanced hysteresis (fig. S17 and table S5) to the spontaneous formation of PbI_2 as a result of aging and enhanced ion migration.

We also tested accelerated aging stability while heating the devices at 85°C in a nitrogen environment. We removed other possible sources of instability such as the organic passivation layer and replaced Spiro-OMeTAD with Poly[bis(4-phenyl)(2,4,6-trimethylphenyl)amine]. We found that the PIRC-targeted devices primarily showed improved stability. Under 500 hours of continuous heating, the device maintained 80% initial performance (Fig. 3F and table S6) whereas the control device only retained 50% of its initial PCE. The PIRC method proposed in this study could enhance both stability and device efficiency which will provide a new direction and push the development of perovskite solar cells; there is also potential for use in other optoelectronic devices such as light-emitting diodes and photodetectors.

REFERENCES AND NOTES

1. N. J. Jeon *et al*., *Nat. Mater.* **13**, 897–903 (2014).
2. D. Bi *et al*., *Nat. Energy* **1**, 16142 (2016).
3. N. J. Jeon *et al*., *Nature* **517**, 476–480 (2015).
4. Y. H. Lin *et al*., *Science* **369**, 96–102 (2020).
5. M. Saliba *et al*., *Science* **354**, 206–209 (2016).
6. H. Zhou *et al*., *Science* **345**, 542–546 (2014).
7. Q. Jiang *et al*., *Nat. Energy* **2**, 16177 (2016).
8. H. Tan *et al*., *Science* **355**, 722–726 (2017).
9. Y. Wang *et al*., *Science* **365**, 687–691 (2019).
10. H. Min *et al*., *Nature* **598**, 444–450 (2021).
11. M. Kim *et al*., *Science* **375**, 302–306 (2022).
12. J. J. Yoo *et al*., *Nature* **590**, 587–593 (2021).
13. X. Zheng *et al*., *Nat. Energy* **2**, 17102 (2017).
14. D. Luo *et al*., *Science* **360**, 1442–1446 (2018).
15. Q. Jiang *et al*., *Nat. Photonics* **13**, 460–466 (2019).
16. R. Wang *et al*., *Science* **366**, 1509–1513 (2019).
17. D. Bi *et al*., *Sci. Adv.* **2**, e1501170 (2016).
18. Q. Jiang *et al*., *Adv. Mater.* **29**, 1703852 (2017).
19. L. Wang, C. McCleese, A. Kovalsky, Y. Zhao, C. Burda, *J. Am. Chem. Soc.* **136**, 12205–12208 (2014).
20. B. W. Park *et al*., *Nat. Commun.* **9**, 3301 (2018).
21. T. J. Jacobsson *et al*., *J. Am. Chem. Soc.* **138**, 10331–10343 (2016).
22. C. J. Tong, L. Li, L. Liu, O. V. Prezhdo, *ACS Energy Lett.* **3**, 1868–1874 (2018).
23. H. Wang *et al*., *Adv. Mater.* **32**, 2000865 (2020).
24. Q. Chen *et al*., *Nano Lett.* **14**, 4158–4163 (2014).
25. F. Liu *et al*., *Adv. Energy Mater.* **6**, 1502206 (2016).
26. J. Holovsky *et al*., *ACS Energy Lett.* **4**, 3011–3017 (2019).
27. M. Kim *et al*., *Joule* **3**, 2179–2192 (2019).
28. R. K. Gunasekaran *et al*., *ChemPhysChem* **19**, 1507–1513 (2018).
29. J. Schoonman, *Chem. Phys. Lett.* **619**, 193–195 (2015).
30. H. Yu, H. Lu, F. Xie, S. Zhou, N. Zhao, *Adv. Funct. Mater.* **26**, 1411–1419 (2016).
31. J. Zhao *et al*., *Sci. Adv.* **3**, eaao5616 (2017).
32. J. M. Azpiroz, E. Mosconi, J. Bisquert, F. D. Angelis, *Energy Environ. Sci.* **8**, 2118–2127 (2015).
33. G. J. Wetzelaer *et al*., *Adv. Mater.* **27**, 1837–1841 (2015).
34. W. H. Nguyen, C. D. Bailie, E. L. Unger, M. D. McGehee, *J. Am. Chem. Soc.* **136**, 10996–11001 (2014).
35. B. Tan *et al*., *Adv. Energy Mater.* **9**, 1901519 (2019).

ACKNOWLEDGMENTS

We thank Y. Yang from the Shanghai Synchrotron Radiation Facility (SSRF), the Shanghai Institute of Applied Physics, Chinese Academy of Sciences for GIWAXIS measurements; we also thank W. Ke from Wuhan University for discussions regarding the PbI_2 complex, and Z. Chu from the Institute of Semiconductors, Chinese Academy of Sciences, for helping with the electroluminescence measurements. **Funding:** J.Y. was supported by National Key Research and Development Program of China (2020YFB1506400) and National Natural Science Foundation of China (61925405, 61634001). H.D. was supported by National Natural Science Foundation of China (61922077) and Y.Y. was supported by Innovation-Driven Project of Central South University (2020CX006). **Author contributions:** J.Y. and Y.Z. conceived the idea; Y.Z. fabricated and characterized the devices; F.M. and Z.Q. participated in device fabrication and preparation for certification; T.S. and H.D. conducted the crystal structure simulation; X.P. and Y.Y. carried out the ion migration activation measurements; F.M., Z.Q., S.Y., X.C., and X.Z. were involved in the data analysis. J.Y. directed and supervised the project, X.Z. cosupervised the project, and J.Y. and Y.Z. cowrote the manuscript. All authors contributed to discussions and finalizing the manuscript. **Competing interests:** A Chinese patent application (CN 202210471109.8) was submitted by the Institute of Semiconductors, Chinese Academy of Sciences, that covers the $(PbI_2)_2RbCl$ used in stabilizing the perovskite. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abp8873
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S6
References (*36*–*45*)
Movie S1

Submitted 4 March 2022; accepted 27 June 2022
10.1126/science.abp8873

CARBON NANOTUBES

DNA-guided lattice remodeling of carbon nanotubes

Zhiwei Lin[1]*†, Leticia C. Beltrán[2], Zeus A. De los Santos[1], Yinong Li[3], Tehseen Adel[4], Jeffrey A Fagan[1], Angela R. Hight Walker[4], Edward H. Egelman[2], Ming Zheng[1]*

Covalent modification of carbon nanotubes is a promising strategy for engineering their electronic structures. However, keeping modification sites in registration with a nanotube lattice is challenging. We report a solution using DNA-directed, guanine (G)-specific cross-linking chemistry. Through DNA screening we identify a sequence, $C_3GC_7GC_3$, whose reaction with an (8,3) enantiomer yields minimum disorder-induced Raman mode intensities and photoluminescence Stokes shift, suggesting ordered defect array formation. Single-particle cryo–electron microscopy shows that the $C_3GC_7GC_3$ functionalized (8,3) has an ordered helical structure with a 6.5 angstroms periodicity. Reaction mechanism analysis suggests that the helical periodicity arises from an array of G-modified carbon-carbon bonds separated by a fixed distance along an armchair helical line. Our findings may be used to remodel nanotube lattices for novel electronic properties.

Synthesis of carbon-based quantum materials can, in principle, employ rich organic chemistry to realize novel properties through atomic-precision structural engineering. More than 50 years ago, W. A. Little proposed a room-temperature organic superconductor model composed of a one-dimensional conducting chain with an array of polarizable side chains attached (*1*). Since then, numerous efforts to verify Little's proposal had been made but all have failed (*2*). In 2016 the mechanism underlying Little's model—electron attraction mediated by polarizable groups—was confirmed for the first time (*3*). The study used a single-wall carbon nanotube (SWCNT) as the one-dimensional conducting chain, along which a nanotube circuitry was constructed to provide a single polarizable "side chain." This work suggests a route to the Little model by chemically implanting polarizable groups in registration with a SWCNT lattice (*4*). However, there exists a major challenge in controlling reaction sites along the nanotube, which seems insurmountable as half the population of carbon atoms on a given SWCNT are chemically equivalent and are enantiomers of the remaining half. We report a DNA-guided chemical reaction to overcome this challenge. We screen reaction products by resonance Raman and photoluminescence (PL) spectroscopy and inspect the structures of promising candidates by single-particle cryo–electron microscopy (cryo-EM). Our findings demonstrate the feasibility of not only creating a wide range of SWCNT derivatives in general but also building a Little model in particular.

Broadly speaking, covalent modification of SWCNTs is a promising route toward organic quantum materials (*5*–*7*). With all atoms on their surfaces, SWCNTs are more amenable than other solid state materials to precision molecular engineering by wet chemistry. In addition, various chiral forms of SWCNTs made available by sorting (*8*) offer a diverse range of electronic structures for further chemical tailoring. Recently, Weisman *et al.* reported a photochemical reaction of DNA-wrapped SWCNTs with singlet oxygen that covalently links guanine (G) to the side wall of SWCNTs (*6*). Even though the chemical nature of the covalent link was not fully revealed and the structure of the reaction product was rather disordered (*9*), we are nevertheless motivated to pursue ordered SWCNT lattice modification with the goal of finding more effective ways to explore nanotube chirality and DNA sequence space.

To speed up sequence screening and to promote ordered structure formation, we have explored conditions for the guanine functionalization reaction of SWCNTs. We find that the previously reported aqueous phase photochemical reaction of Rose Bengal (RB)–mediated guanine cross-linking with SWCNTs (*6*) also works in mixed solvents containing both methanol and water. In 50% v/v methanol, the reaction proceeds slower (60 min) than that in water (15 min); further, consumption of RB sensitizer is decreased considerably and RB binding to SWCNTs—which may adversely affect DNA wrapping structures and cross-linking sites—is also minimized. All these changes should favor homogeneous product formation. Combining this new reaction condition with a previously established process (*10*) for DNA-surfactant exchange, we have devised a one-pot chemistry (Fig. 1A and figs. S1 to S5) that can efficiently react any G-containing sequence with any single-chirality SWCNT species [e.g., (6,5), (9,1), (8,3), and so on] purified independently through various techniques (*8*). Excess DNA is present during the reaction to refill exposed nanotube surfaces arising from reaction-induced DNA structure contraction, eliminating a potential source of inhomogeneous functionalization.

We have employed spectroscopic tools to monitor the reaction (Fig. 1 and figs. S6 and S7). Figures 1B to 1D present a data set from a reaction on an (8,3) enantiomer (normalized circular dichroism signal at the E_{22} = – 42 mdeg, which according to theoretical analysis corresponds to a right-handed enantiomer) (*11*). Hereafter we denote the enantiomer simply as (8,3) unless indicated otherwise. Figures 1B and 1C respectively show the expected red shift of the E_{11} absorption and PL peak after functionalization. Figure 1D shows resonance Raman spectra of unfunctionalized (black trace) and functionalized (8,3) (red trace). In addition to the well-documented D peak, we find another disorder-induced peak: the intermediate frequency mode (IFM) (*12*, *13*) at 387 cm^{-1}, with overtone 2 IFM at 775 cm^{-1} and combination modes with D: D ± IFM at 905 and 1680 cm^{-1}, respectively. There is considerable intensity enhancement of the D and IFM modes in the functionalized tubes. By contrast, 2IFM and D ± IFM peaks remain weak and unchanged after functionalization, consistent with their origin from two-phonon, second-order scattering processes (*12*).

To gain insight into the reaction mechanism, we have functionalized (8,3) using a set of DNA sequences with varying G content. With increasing G content the functionalized (8,3) shows a gradual increase in both D and IFM peak intensity (Fig. 2, A and B) in addition to more red-shifted and broadened PL and absorbance peaks (Fig. 2, C and D). These observations are consistent with a previous study which used mixed chirality tubes (*6*). However, using chirality-pure SWCNTs allows us to unambiguously determine the absorption and PL peak positions of the functionalized tube as well as calculate its absorption peak shift and Stokes shift (SS), i.e., the energy difference between the absorption and PL peak, as a function of G content or defect density. The absorption peak shift increases with G content (Fig. 2E), but its amplitude is <27 meV, about ten times as small as that typically observed for an sp^3 defect (*14*). This difference is notable considering that the defect density (≈10°/nm) in guanine-modified tubes (*15*) is about two orders of magnitude higher than that in sp^3-modified tubes (*5*). The SS also increases with defect density (Fig. 2E), a trend opposite that observed for sp^3 defects (*14*). In addition, we find a quantitative relationship between SS and full width at half maximum

[1]Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [2]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA. [3]South China Advanced Institute for Soft Matter Science and Technology, School of Emergent Soft Matter, South China University of Technology, Guangzhou 510640, China. [4]Quantum Metrology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.
***Corresponding author. Email: zhiweilin@scut.edu.cn (Z.L.); ming.zheng@nist.gov (M.Z.)**
†Present address: South China Advanced Institute for Soft Matter Science and Technology, School of Emergent Soft Matter, South China University of Technology, Guangzhou 510640, China.

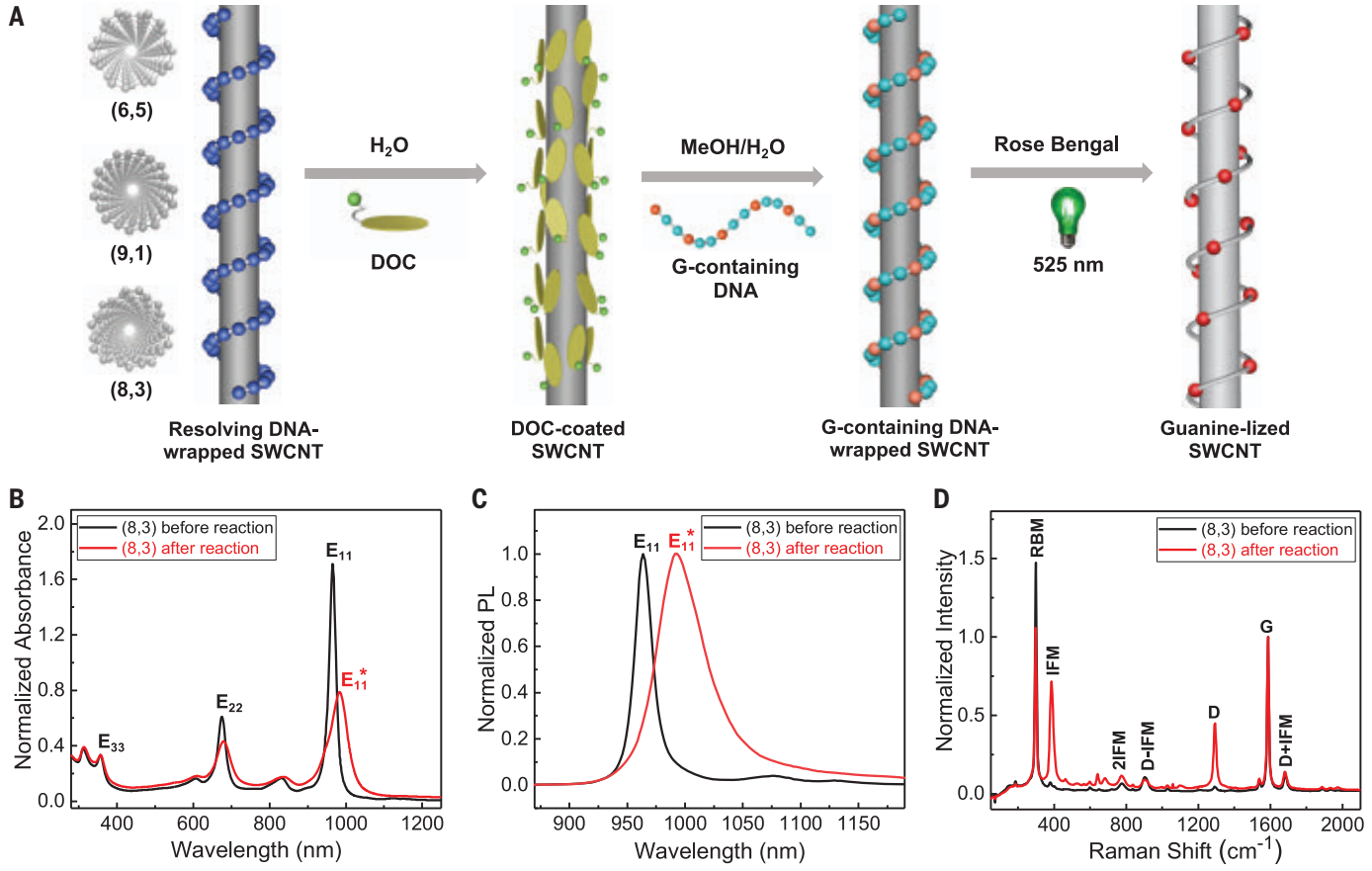


Fig. 1. One-pot photochemical reaction scheme and spectral characterization. (**A**) Reaction scheme starting from resolving DNA-wrapped to guanine-functionalized (guanine-lized) SWCNT carried out in a single pot. DOC, sodium deoxycholate. (**B**) Absorption (normalized at E_{33}). (**C**) PL (with peak intensity normalized to 1). (**D**) Resonance Raman spectra (normalized at the "G" peak) of (8,3) before and after the reaction. Spectra in (C) and (**D**) are measured with 671 nm excitation corresponding to E_{22} of (8,3). The DNA sequence used in this experiment is $(GCC)_{12}$. See fig. S8 for data shown in (B) and (C) before normalization.

(W) of the PL peaks, suggesting that modified and unmodified carbon atoms are isovalent as discussed below. In Fig. 2F we plot SS versus W^2 and fit the data well (R^2 = 0.98) with T = 349.3 K using Eq. 1, where k is the Boltzmann constant and T is the effective exciton temperature equal to or above the ambient temperature.

$$SS = \frac{W^2}{8 \ln 2\, kT} \cong 0.18 \frac{W^2}{kT} \qquad (1)$$

Eq. 1 has been used to describe excitons in two-dimensional (2D) quantum wells and 3D alloy semiconductors in which disordered isovalent substitution creates shallow traps (*16*, *17*). It attributes the observed SS to the thermalization of excitons in an inhomogeneously broadened band. Combining all of these spectroscopic observations, we exclude sp^3 defect formation by guanine functionalization and conclude that the chemistry employed creates a modified sp^2 defect that is isovalent to the original sp^2 carbon in the pristine SWCNT.

To differentiate the two types of disorders originating from the defect density itself and defect distribution pattern, we have designed a set of 15-mer G/C sequences containing two Gs separated by a varying number of Cs (Fig. 3A) for the (8,3) functionalization chemistry. We find that disorder-induced D and IFM peak intensities change as a function of inter-G spacing, reaching an unexpectedly deep minimum when the spacing is 7 (Fig. 3, B and C). The SS and PL peak linewidth also show minimum values at that spacing (Fig. 3, D and E). These observations appear to be dependent on nanotube chirality as left- and right-handed (6,5)s functionalized by the same set of sequences yield different spectral patterns (fig. S7).

We offer a qualitative analysis of the data shown in Fig. 3. According to the mechanism of defect-induced Raman modes (*12*, *18*), the observed peak intensity is proportional to the extent of elastic electron scattering by defects. For the sequences used in our study we estimate that a SWCNT contains 1 to 5 guanine-modified sites per nanometer tube length based on molecular dynamics simulations of a typical DNA wrapping structure on a SWCNT (*15*). Because excitons are 2 to 3 nm in size (*19*) electron scattering is expected to involve multiple defect sites and is sensitive to not only the defect density but also the degree of order in the defect array. The minimum D and IFM peak intensities shown in Fig. 3 are thus interpreted as resulting from an ordered defect array generated by the 2G-7 sequence. This conclusion is also consistent with the observed minimum SS for the same sequence, as SS is another measure of disorder for semiconductors with isovalent substitution (*16*, *17*).

To independently evaluate the spectroscopy-derived result, we have applied single-particle cryo-EM to measure a 2G-7-(8,3) hybrid structure before and after functionalization. Cryo-EM imaging of 2G-7 functionalized (8,3) reveals filaments ~20 Å in diameter (fig. S11). In the nonfunctionalized structure coated with 2G-7, an averaged power spectrum from ≈ 2×10^5 images of particle segments yields no detectable features (fig. S12), implying a disordered DNA wrapping structure. In the functionalized structure we have detected a layer line pattern characteristic of a one-start helix with a 6.5 Å helical pitch visible from the averaged power spectrum of selected segments (Fig. 4A). The

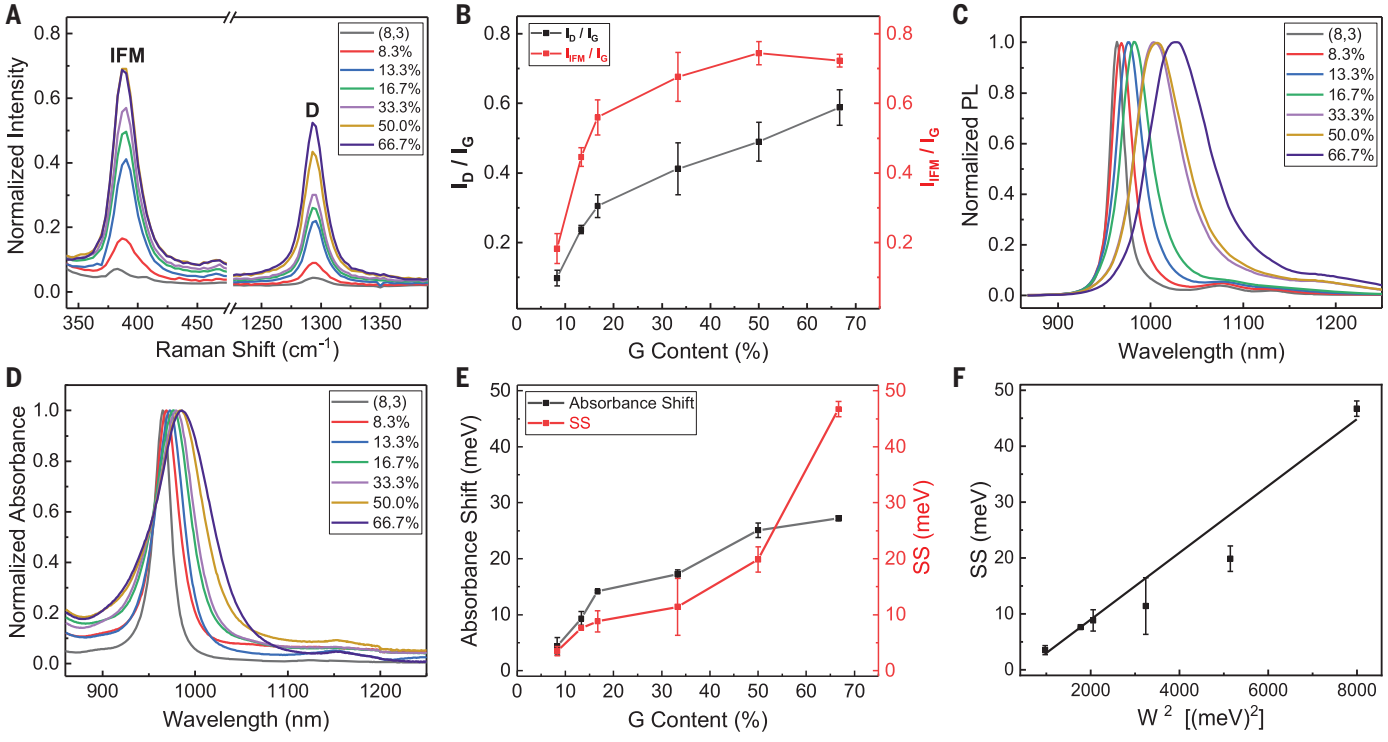


Fig. 2. Spectroscopic characterization of (8,3) functionalized with DNA of varying G content (see table S1 for sequence information). (**A** and **B**) IFM and D peak intensity (normalized by the intensity of the "G" peak) profile. (**C** and **D**) PL and absorbance profile. (**E**) Absorbance shift and SS profile. (**F**) SS versus W^2 and a linear fit (R^2 = 0.98) using Eq. 1 with T = 349.3 K. Error bars shown here and elsewhere in this work are derived from three independently measured values for each of three independently prepared samples. Spectra in (A) and (C), are measured with 671 nm excitation. See fig. S9 for original data.

helical structure is also visible in the 2D class average (Fig. 4B). Ab initio reconstruction—an unbiased and reference-free approach—was employed to generate a 3D reconstruction of the DNA-coated nanotube (Fig. 4C). Figure 4D shows a highly averaged version of that reconstruction, with an atomic model for the SWCNT. Because of the low signal-to-noise ratio in these cryo-EM images and the apparent lack of any other periodicities present, we have imposed a helical averaging of the density along the 6.5 Å pitch helix. The unfiltered 3D map (Fig. 4C) shows a coherence length of ~10 nm, shorter than the full length of the segments used. The density in Fig. 4, C and D, is shown as a left-handed helix although true handedness was not determined.

The 6.5 Å periodicity observed by cryo-EM provides support for the model of ordered defect array formation and is also a valuable clue regarding the mechanism of guanine functionalization. As discussed earlier, a correct reaction mechanism should yield a modified carbon that largely maintains its sp^2 character. Indeed, such chemistry has been well documented (*20*, *21*) and forms the basis for our proposal as shown in Fig. 4, E and F, and fig. S14B. Guanine oxidation by singlet oxygen has a plethora of pathways and products depending on reaction conditions and structural context but a common initial step is oxidation of the C_8 carbon on the imidazole ring (*22*). We propose that the C_8 carbon becomes electrophilic upon oxidation and then undergoes a 2+1 cycloaddition with a nearby C-C bond on a SWCNT to yield a three-membered ring. This is followed by C-C bond cleavage and concomitant ring opening as a result of ring strain, leaving C_8 to bridge the two carbons from the SWCNT and restore their sp^2 character. Theoretical calculations (*20*) predict that this type of reaction is most favorable on C-C bonds with large curvatures or bond strains, consistent with a previous observation from the guanine functionalization chemistry [figure 2C in (*6*)]. In (8,3) there are three types of C-C bonds with distinct curvatures; of these, C-C bonds along the helical armchair line shown in Fig. 4E possess the largest curvature. We note that the pitch (p) of this helical line has an intrinsic length scale of (8,3) determined solely by its chiral index (n, m) and C-C bond length $a_c : p = a_c\sqrt{n^2 + m^2 + nm}\ \frac{n-m}{n+m} = 6.45$ Å (for $a_c = 1.44$ Å), matching that observed by cryo-EM. We therefore propose a 2G-7 wrapping structure in which each of the 2Gs is covalently linked to a C-C bond along the armchair line, resulting in pinning of the DNA backbone along the same armchair line. Consistent with this functionalization-induced DNA pinning is our observation that 2G-5 and 2G-6 functionalized (8,3) also exhibit the same 6.5 Å helical pitch (fig. S13). Manual model building for 2G-7 functionalized (8,3) followed by energy minimization yields a model as shown in fig. S14A, in which two adjacent G modification sites are separated by five C-C bonds along the armchair line. This equal G spacing explains the minimum spectroscopy-derived structure disorder for 2G-7 functionalized (8,3).

In summary, we show that ordered SWCNT modification can be achieved by taking advantage of DNA sequence control over the spacing between adjacent reaction sites, and SWCNT's bond curvature–dependent reactivity. Our findings demonstrate the chemical feasibility of building a Little model. DNA-guided remodeling breaks the original symmetry of the nanotube lattice and therefore should lead to new modes of low-energy electronic excitation. Theoretical analysis shows that helical modification of a SWCNT in registry with its lattice may induce topological electronic behavior (*23*, *24*), suggesting that the chemistry we report here might be used to explore topological physics. We envision

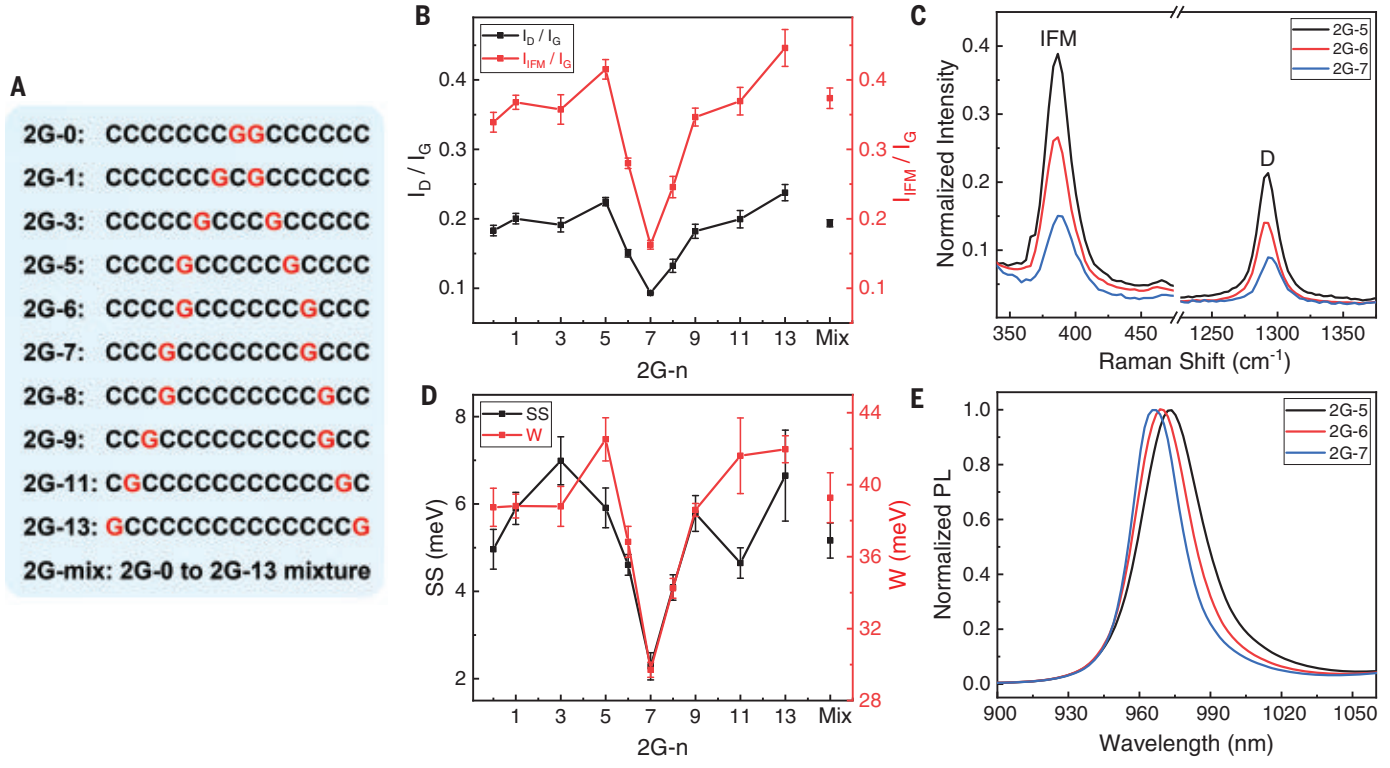


Fig. 3. Screening DNA sequences for the ordered defect array. (**A**) DNA sequence used for the screening. (**B**) D and IFM peak intensity profile. (**C**) Raman spectra of 2G-5, 2G-6, and 2G-7 functionalized (8,3) (normalized by the "G" peak). (**D**) SS and *W* profile. (**E**) PL spectra of the three samples. Spectra in (C) and (E) are measured with 671 nm excitation. See fig. S10 for original data.

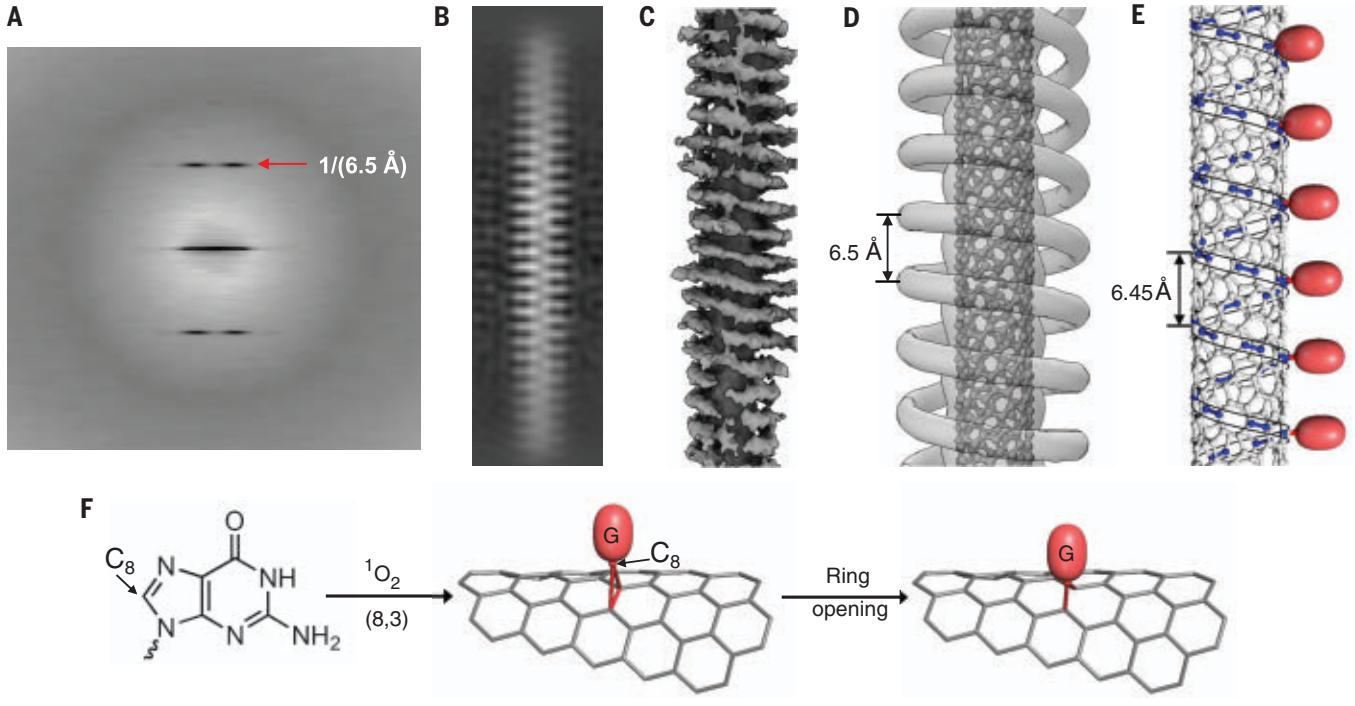


Fig. 4. Cryo-EM–derived structure model for 2G-7 functionalized (8,3) and reaction mechanism. (**A**) Averaged power spectrum from ~44,000 particles. Red arrow pointing to a layer line with a spacing of 1/(6.5 Å) from the equator. (**B**) Image of 2D class average. (**C**) Low-resolution 3D map generated by unbiased, reference-free approach displaying coherence that extends over an axial length of ≈ 10 nm. (**D**) 3D map corresponding to averaged density along the 6.5 Å pitch helix. (**E**) Carbon-carbon bonds that have maximum bond curvature, highlighted in blue along an armchair helical line of (8,3), along which every sixth carbon-carbon bond is modified by a guanine (red balls) according to modeling (fig. S14A). (**F**) A proposed reaction mechanism.

that future work will introduce diversity of functional groups, increase the coherent length of ordered modification, and eventually enable discovery of organic quantum materials and carbon-based, metal-free catalysts (*25*, *26*).

REFERENCES AND NOTES

1. W. A. Little, *Phys. Rev.* **134**, A1416–A1424 (1964).
2. D. Jérome, *Science* **252**, 1509–1514 (1991).
3. A. Hamo *et al*., *Nature* **535**, 395–400 (2016).
4. T. Kontos, *Nature* **535**, 362–363 (2016).
5. Y. Piao *et al*., *Nat. Chem.* **5**, 840–845 (2013).
6. Y. Zheng, S. M. Bachilo, R. B. Weisman, *ACS Nano* **13**, 8222–8228 (2019).
7. F. Yang *et al*., *Chem. Rev.* **120**, 2693–2758 (2020).
8. M. Zheng, *Top. Curr. Chem. (Cham)* **375**, 13 (2017).
9. Y. Zheng *et al*., *ACS Nano* **15**, 10406–10414 (2021).
10. J. K. Streit, J. A. Fagan, M. Zheng, *Anal. Chem.* **89**, 10496–10503 (2017).
11. N. Sato, Y. Tatsumi, R. Saito, *Phys. Rev. B* **95**, 155436 (2017).
12. A. Vierck, F. Gannott, M. Schweiger, J. Zaumseil, J. Maultzsch, *Carbon* **117**, 360–366 (2017).
13. T. Inaba, Y. Tanaka, S. Konabe, Y. Homma, *J. Phys. Chem. C* **122**, 9184–9190 (2018).
14. F. J. Berger *et al*., *ACS Nano* **13**, 9259–9269 (2019).
15. D. Roxbury, A. Jagota, J. Mittal, *J. Phys. Chem. B* **117**, 132–140 (2013).
16. M. Gurioli, A. Vinattieri, J. Martinez-Pastor, M. Colocci, *Phys. Rev. B Condens. Matter* **50**, 11817–11826 (1994).
17. M. Felici *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **74**, 085203 (2006).
18. C. Thomsen, S. Reich, *Phys. Rev. Lett.* **85**, 5214–5217 (2000).
19. L. Lüer *et al*., *Nat. Phys.* **5**, 54–58 (2009).
20. Y.-S. Lee, N. Marzari, *Phys. Rev. Lett.* **97**, 116801 (2006).
21. A. Setaro *et al*., *Nat. Commun.* **8**, 14281 (2017).
22. Y. Ye *et al*., *J. Am. Chem. Soc.* **125**, 13926–13927 (2003).
23. V. I. Puller, S. V. Rotkin, *Europhys. Lett.* **77**, 27006 (2007) (EPL).
24. C. Hu, V. Michaud-Rioux, W. Yao, H. Guo, *Nano Lett.* **19**, 4146–4150 (2019).
25. K. Gong, F. Du, Z. Xia, M. Durstock, L. Dai, *Science* **323**, 760–764 (2009).
26. X. Liu, L. Dai, *Nat. Rev. Mater.* **1**, 16064 (2016).

ACKNOWLEDGMENTS

We thank B. Weisman and S. Tretiak for useful discussions. **Funding:** This work was supported by NIST internal funding and NIH GM122510 (to E.H.E.). Z.A.D. acknowledges an NRC postdoctoral fellowship. **Author contributions:** Z.L. and M.Z. conceived the idea. Z.L. and Z.A.D. performed the experiments, J.A.F. provided the samples, L.C.B. and E.H.E. performed cryo-EM and analysis, Z.L. and Y.L. performed the molecular modeling, T.A. and A.R.H.W. performed the Raman analysis, and Z.L., M.Z., L.C.B., and E.H.E. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4628
Materials and Methods
Figs. S1 to S14
Table S1
References (*27*–*30*)

Submitted 5 February 2022; accepted 6 June 2022
10.1126/science.abo4628

SOLID-STATE IONICS

Nanosecond protonic programmable resistors for analog deep learning

Murat Onen[1,2]*, Nicolas Emond[2,3], Baoming Wang[2,3], Difei Zhang[1,2], Frances M. Ross[2,3], Ju Li[2,3,4]*, Bilge Yildiz[2,3,4]*, Jesús A. del Alamo[1,2]*

Nanoscale ionic programmable resistors for analog deep learning are 1000 times smaller than biological cells, but it is not yet clear how much faster they can be relative to neurons and synapses. Scaling analyses of ionic transport and charge-transfer reaction rates point to operation in the nonlinear regime, where extreme electric fields are present within the solid electrolyte and its interfaces. In this work, we generated silicon-compatible nanoscale protonic programmable resistors with highly desirable characteristics under extreme electric fields. This operation regime enabled controlled shuttling and intercalation of protons in nanoseconds at room temperature in an energy-efficient manner. The devices showed symmetric, linear, and reversible modulation characteristics with many conductance states covering a 20× dynamic range. Thus, the space-time-energy performance of the all–solid-state artificial synapses can greatly exceed that of their biological counterparts.

Aqueous ionics with mobile Na^+, K^+, Ca^{2+}, and other ions underpin biological information processing. In neurons and synapses, the action potential of magnitude ~100 mV evolves over a characteristic time scale of milliseconds, which fundamentally limits the speed of thinking and reflexes of animals. Because liquid water decomposes at voltages >1.23 V, a weak action potential is understandable. However, with human-made solid-state neurons and synapses (*1*–*3*) for analog machine learning, research efforts are no longer limited by the stability window of aqueous electrolyte. Furthermore, it is also possible to fabricate devices that are much smaller than biological neurons, by a length-scale shrinkage factor of 10^3 (from ~10 μm to ~10 nm). This possibility raises a fundamental question about how much faster we can train such "artificial synapses"—i.e., what is the ultimate speed limit to solid-state ionics–based analog deep learning? Also, when we approach that speed limit, how energy intensive would the training be, given that dissipative processes generally produce more entropy per task the faster that task is accomplished?

Fundamental ionics arguments seem to call for high voltage and small length scales—that is, an extreme programming field approach (*4*–*10*). Transport of ions (such as H^+) inside a solid electrolyte (SE) layer and a mixed ionic-electronic conductor (MIEC) conductance channel layer, as well as charge-transfer reactions at the SE/MIEC interfaces, scale monotonically and nonlinearly with the applied voltage. Therefore the speed of change in the nonvolatile state of artificial neurons should increase with increasing electric field. However, there is a dielectric breakdown limit (*11*) to solids, as well as a thermodynamic electrochemical stability window for SEs. Therefore, the programming field should be raised as high as possible, but not so high as to permanently damage the SE. In this work, we demonstrate that such an approach produced exceedingly fast nanoionic devices (5 ns), at least 10^4 times as fast as biological synapses, when the voltage was on the order of +10/−8.5 V, yielding an extremely high electric field of ~1 V/nm across the proton-conducting phosphosilicate glass (PSG) SE. Surprisingly, we also show that under appropriate operating parameters, this nonlinear ionic-electronic device was robust and reversible, operating successfully over millions of cycles. Although the layout of our three-terminal device is analogous to that of a solid-state battery, it operated 10^{10} times as fast as conventional solid-state batteries. Finally, even when approaching the ultimate speed limit of our devices, the heat generated per programming task was still favorable compared with that generated by a human synapse (~10 fJ per state).

Interest in engineering the ideal programmable resistor for analog computing applications has skyrocketed owing to increasing workloads of deep learning problems (*12*–*14*) and diminishing expectations for transistor performance improvements through size scaling (*15*). Highly optimized digital application–specific integrated circuits that run reduced precision arithmetic operations can still manage inference tasks (*16*); however, the resolution requirements for training tasks do not allow

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. [2]MIT-IBM Watson AI Lab, 75 Binney St., Cambridge, MA 02142, USA. [3]Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. [4]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA.
***Corresponding author. Email: monen@mit.edu (M.O.); alamo@mit.edu (J.A.d.A.); byildiz@mit.edu (B.Y.); liju@mit.edu (J.L.)**

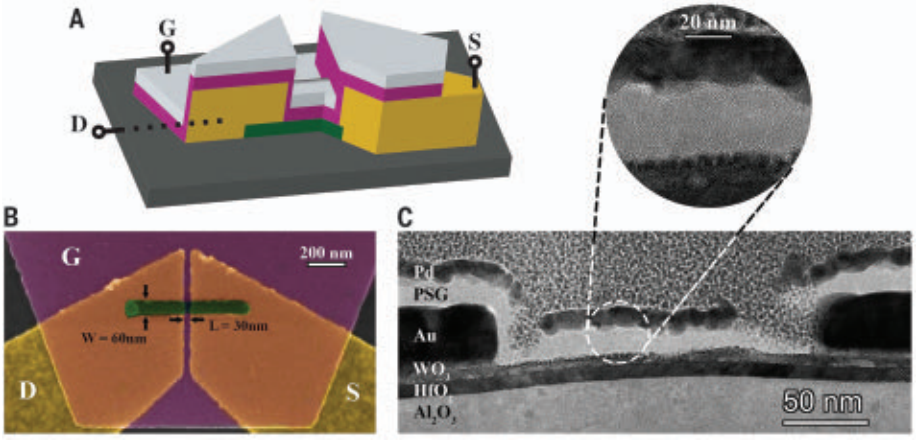


Fig. 1. Nanosecond protonic programmable resistors. (**A**) Three-dimensional illustration of the protonic programmable resistors studied in this work. Au (yellow), WO_3 (green), PSG (magenta), and Pd (gray) layers are indicated. As a result of an engineered sidewall, the Pd layer that overlaps with the channel electrodes is isolated from the remainder of the gate electrode. G, gate; D, drain; S, source. (**B**) False-colored top-view SEM image of a fabricated device with a 30-nm–by–60-nm channel. (**C**) Transmission electron microscopy (TEM) cross-sectional image of a protonic programmable resistor that had previously been extensively modulated in the ultrafast regime.

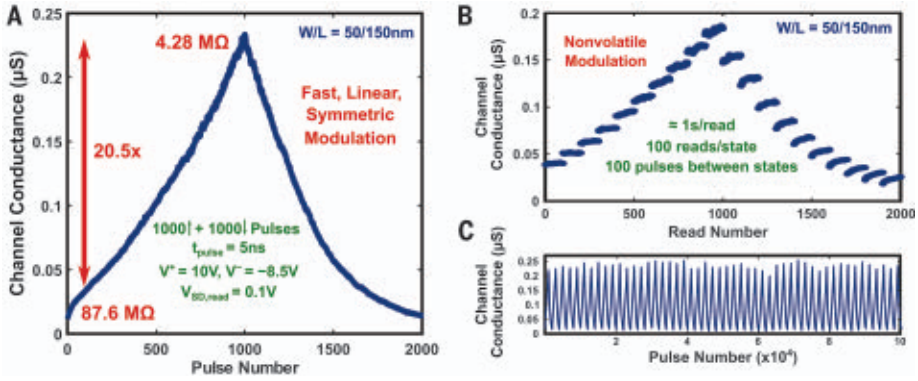


Fig. 2. Ultrafast and energy-efficient modulation characteristics of protonic programmable resistors. (**A**) Modulation performance of a 50-nm–by–150-nm protonic device with 10-nm PSG, showing fast (5 ns per pulse), nearly linear, and symmetric characteristics. W, width; L, length. (**B**) Retention behavior of the protonic device for ≈100 s at different conductance levels over the full dynamic range. (**C**) Endurance characterization of the protonic device, displaying nondegrading modulation over 10^5 pulses conducted over 30 hours.

sufficient bit reduction (*17*) and make beyond-digital approaches imperative. The core idea behind analog training accelerators is to process information locally using physical device properties instead of conventional Boolean arithmetic—i.e., using Ohm's and Kirchhoff's laws for the matrix inner product (*18*) and threshold-based updating for the outer product (*19*). However, the performance benefits attained by analog training processors are dependent on a set of very strict properties: The device must be fast, energy-efficient, nonvolatile, and reversible, and it must show symmetric conductance modulation with many conductance states across a large dynamic range (*20*, *21*). Unfortunately, readily available memory technologies [e.g., phase-changing (*22*, *23*), filamentary (*24*), bulk-switching (*25*–*28*), and ferroelectric (*29*) resistive memories] do not completely satisfy these requirements. As a result, demonstrations so far have required additional control circuitry and serial operations (*22*–*24*) to correct such imperfections, at the expense of substantially diminished space-time-energy performance.

Because the core idea of neural network training is to tune the state parameters (weights) through many small incremental modifications (*12*, *13*), it is reasonable that devices originally designed for information storage purposes (i.e., memory technologies) do not perform well in this information processing application. Devices for analog deep learning (i.e., training) must be primarily optimized for state transition (i.e., modulation) rather than long-term state preservation (i.e., nonvolatility). The retention of the weight parameters is obviously important such that the information learned from previous inputs is not lost faster than the ongoing optimization. However, given that the purpose of the network is not to memorize correct labels for the training dataset but to find generalized features instead, retaining exact values for extended durations of time loses relevance. The best example for this argument can be found in digital neural network training, for which it is a common practice to simulate weight decay on values that could otherwise be retained for indefinite time (*30*). Because the properties that ensure long retention times contradict those that enable high-speed and low-energy modulation (*31*), we conclude that for training implementations the focus should be on the latter two, trading retention down to the ~100-s–to–1000-s range for nanosecond and subpicojoule performance [see the supplementary materials (SM)].

Any memristive technology that aims to realize practical analog processors must contain a high density of devices built on top of standard complementary metal-oxide semiconductor (CMOS) logic and therefore must rely on CMOS-compatible materials and involve back-end-of-line–compatible processes. Moreover, device operation should not depend on unconventional environmental conditions such as high temperature or humidity. With these concepts in mind, we developed an integration-friendly technology with inorganic solid materials that are native to conventional Si processing: WO_3 as the active channel material, nanoporous PSG as the protonic SE, and Pd as the hydrogen reservoir and controlling gate. The basic operating principle of the device relies on modulating the channel conductance via the electrochemically controlled intercalation of protons (chosen as the ion with the smallest radius and lightest mass) into WO_3 (*32*, *33*).

The rationale for choosing these materials is as follows: WO_3 is known for its conductance modulation and electrochromism upon ion intercalation (*25*, *32*–*34*) with well-established dynamics (*32*, *35*). For analog deep learning applications, the channel material must also have high base resistivity (*19*), proton insertion rate, and sensitivity. We found that superior modulation characteristics result from reactive sputtering of WO_3 at room temperature, followed by a 400°C annealing process that both oxidizes and crystallizes the material (see SM for process details). Regarding the electrolyte, the material must simultaneously show very high electronic resistivity and high proton conductivity. In our previous work, we identified nanoporous PSG as an outstanding SE (*36*–*38*) that displays both properties at room temperature without the need for hydration (i.e., humidity) (*33*). Finally, Pd was chosen as the gate metal and hydrogen reservoir (PdH_x), owing to its capability to take up and store hydrogen (*39*).

The three-terminal protonic programmable resistors studied in this work (Fig. 1A) were

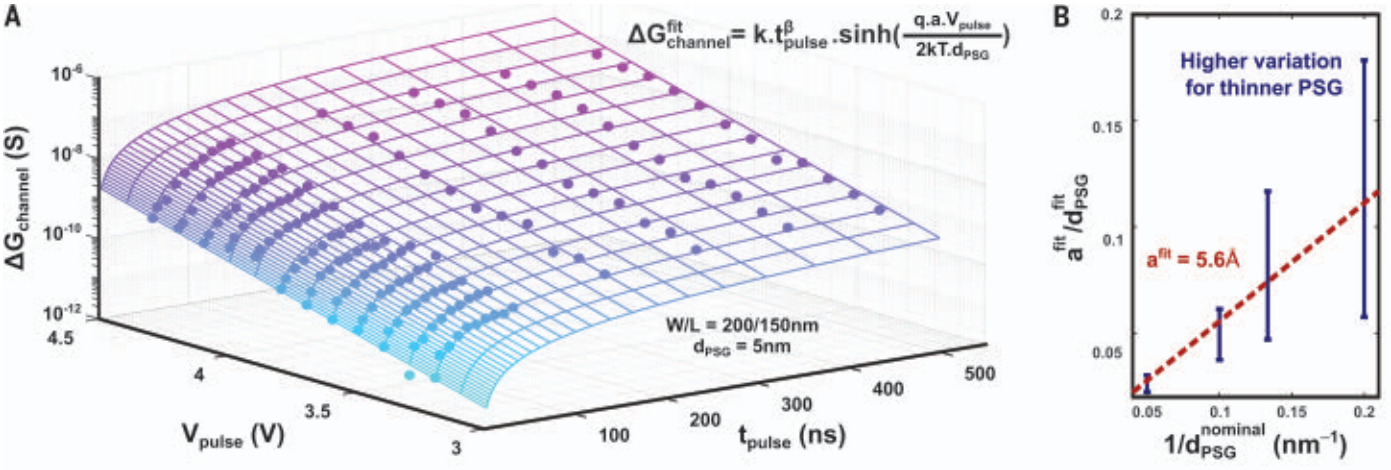


Fig. 3. Voltage dependence of conductance modulation. (**A**) Channel conductance change per pulse ($\Delta G_{channel}$) for different pairs of pulse amplitude (V_{pulse}) and duration (t_{pulse}) for a device with 5-nm PSG thickness. Points represent averaged experimental data over 1000 identical pulses for a given V_{pulse}-t_{pulse} pair, whereas the meshed surface represents the fitted result. The apparent hopping distance (a) and power factor (β) values used to fit the data are 5.6 Å and 1.2, respectively. The pulse time dependence of the protonation-deprotonation dynamics was empirically approximated as a power law ($\propto t_{pulse}^{1.2}$). (**B**) Estimation of the apparent hopping distance from repeating the same experiment detailed in Fig. 3A for 25 devices selected over four different chips with different PSG thicknesses. Error bars indicate the SD of values acquired from different devices with the same PSG thickness.

fabricated through a series of electron-beam lithography processes. A self-aligned gate structure was used to scale down device dimensions, avoiding mask alignment limitations. In a key design aspect of this process, the Pd layer was overlaid across a large region (Fig. 1B), which was then used as an etch mask for the PSG layer beneath. The height of the channel electrodes (Au) was calibrated such that the PSG layer could cover the sidewalls and the Pd layer that overlapped with the channel electrodes was disconnected from the gate electrode (Fig. 1C), avoiding the charging of those regions and unnecessary capacitance. This configuration was intended to maximize energy efficiency. The complete fabrication flow, additional metrology results, and yield details can be found in the SM.

Before testing, the devices were exposed to forming gas (3% H_2 in N_2) at room temperature for protonation of the Pd reservoir and then pumped down to vacuum. Future iterations of these protonic devices will be encapsulated to avoid environmental preconditioning steps. Figure 2A shows the channel conductance modulation of a 50-nm–by–150-nm device with 10-nm-thick PSG for 1000 protonating voltage pulses (V^+ = 10 V) followed by 1000 deprotonating pulses (V^- = −8.5 V). Between successive pulses, the channel conductance was read under drain-source voltage (V_{DS}) = 0.1 V and gate current (I_G) = 0 conditions and averaged for ~1 s. The devices displayed nearly ideal characteristics in terms of (i) high modulation speed, responding to 5-ns voltage pulses; (ii) nearly linear and symmetric behavior for incremental and decremental changes; (iii) conductance retention characteristics over durations longer than ~10^{10} times the unit pulse time (Fig. 2B); (iv) dynamic conductance range of 20×; (v) optimal base resistance of 88 megohms for readout (*19*, *32*); and (vi) preservation of these ideal properties without any degradation over extended time and use (Fig. 2C).

Moreover, the devices showed excellent energy efficiency under this ultrafast operation, the gate current supplied during each pulse being too small to be precisely measured for the small devices. The energy consumption during the transients was estimated to be ~2.5 fJ per pulse (see SM), which is a technology-agnostic overhead related to charging and discharging the gate capacitance. On the other hand, the energy consumed in proton transfer while the 5-ns voltage pulse was at its peak value was estimated to be ~15 aJ per pulse for the device whose performance is shown in Fig. 2. This value was based on a dc gate current measurement of ~30 nA at 10 V for a 750-nm–by–1000-nm device. This latter value is associated with the efficient shuttling of ions within the gate stack under the high electric field. Additional data and calculations related to the energy consumption can be found in the SM.

As a result, the combined material, processing, and performance benefits of the all–solid-state protonic devices demonstrated here exceed those of nonvolatile memory technologies.

To explain these desirable modulation characteristics, we have developed a model for device operation that consists of two key parts: (i) proton transport in the PSG and (ii) proton-coupled electron-transfer reaction rates at the PSG/electrode interfaces. Both processes have qualitatively similar formalisms and dependencies—the former is governed by hopping conduction in a disordered solid with random site energies (*4*–*7*, *10*), whereas the latter is determined by the Butler-Volmer charge-transfer kinetics (*8*, *9*). Conventionally, a high activation energy for protons to be freed from their sites (e.g., Si–O–H or P–O–H) results in low ionic conductivity (*40*) and ultimately limits the operation speed at room temperature. For amorphous SiO_2, this activation energy was reported to be 0.38 eV, which is much larger than the thermal energy in ambient conditions (25.9 meV) (*41*).

In the presence of a high electric field, the energy barrier to ion conduction is lowered in the field direction, yielding an enhanced proton hopping current

$$I_G(V_{pulse}) \propto \sinh\left(\frac{qaV_{pulse}}{2k_BTd_{PSG}}\right) \quad (1)$$

where q is the electron charge, k_B is the Boltzmann constant, T is the temperature, a is the apparent hopping distance, and d_{PSG} is the thickness of the electrolyte (*5*). Owing to the high resistance of the PSG layer, all of the pulse voltage is assumed to drop across the electrolyte ($V_{PSG} \sim V_{pulse}$; see SM for details). Figure 3A shows the experimentally observed conductance change per pulse ($\Delta G_{channel}$) as a function of pulse amplitude (V_{pulse}) and pulse time (t_{pulse}) for a device with d_{PSG} = 5 nm. Over a range of electric fields similar to that detailed in Fig. 2, the results shown in Fig. 3 closely followed Eq. 1, as indicated by the lines. Furthermore, experiments for devices with different d_{PSG} values allowed us to extract a hopping distance for the PSG of 5.6 Å (Fig. 3B). This result is in good agreement with previously reported values for amorphous silica glasses (*41*–*43*).

In our case, the electric field (≈1 V/nm) across the hopping distance ($a \approx 0.5$ nm) was so high that it might completely remove the

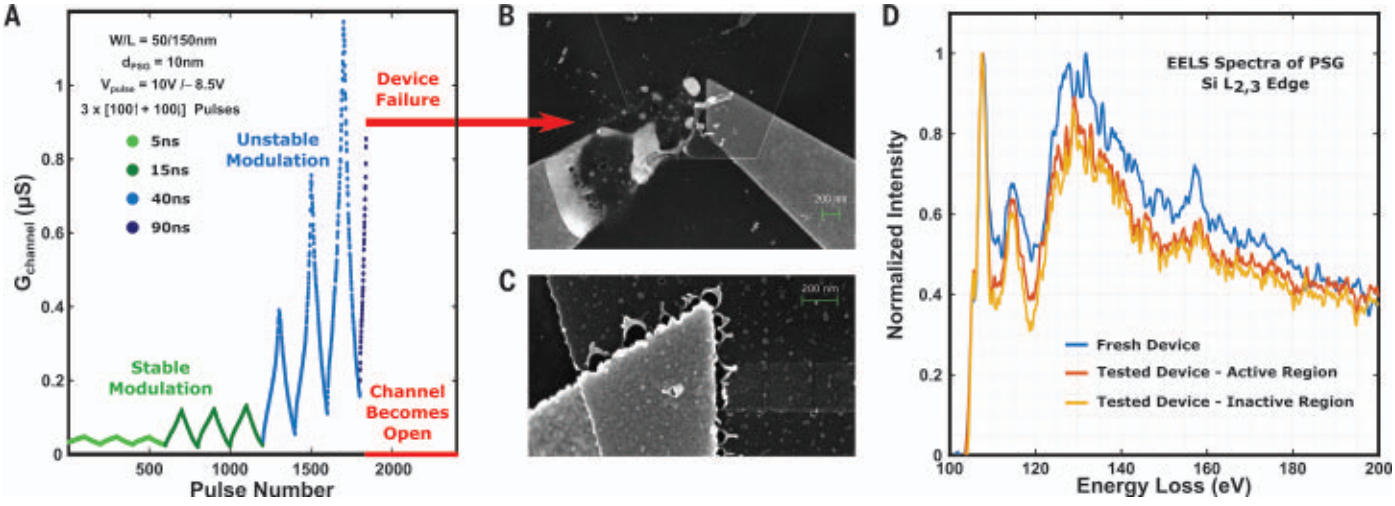


Fig. 4. Modulation dynamics for short and long pulse durations. (**A**) Channel conductance modulation of the device whose performance is presented in Fig. 2 for increasing pulse duration under the same programming voltages. Between pulses, the channel conductance is read and averaged for ≈1 s. (**B**) SEM image of the device after the experiment shown in (A). (**C**) SEM image of another device captured during early degradation. (**D**) Si-$L_{2,3}$ EELS spectra of the PSG layer for tested (red and orange) and fresh (blue) devices. The testing conditions of the device were identical to those given in Fig. 2, operating under the ultrafast regime only.

activation barrier within the PSG (≈0.4 eV). This effect resolves the bottleneck of low proton conduction at room temperature, thus enabling high-speed operation. Such an effect was predicted in simulations (*44*) but has not been previously observed experimentally, as the required conditions are beyond the breakdown field or the electrochemical stability window of traditional electrolyte materials (*45*). Instead, PSG allows a high critical field (8 to 15 MV/cm) (*46*, *47*) in addition to a moderate base proton conductivity (*36*–*38*), making it an ideal electrolyte choice for this application. Moreover, under the classical statistical mechanics scenario, reducing the migration energy below a few k_BT boosts the likelihood of classical or quantum ballistic motion of protons. Indeed, in liquid water, proton motion is known to have a pronounced quantum character, with activationless quantum nuclear dynamics in some exchange events between water molecules (*48*). Thus, under such high electric fields, "quantum ballistic" transport of protons in solids may be realized.

Unlike the conductive filament devices controlled by angstrom-scale short-range ion motion at the tip of the filament (*49*), our protonic devices rely on long-range (~10 nm) and uniform partitioning of protons across the electrolyte with the protons then behaving as dynamic dopants in the entire H_yWO_3 channel (*32*, *35*). Furthermore, conducting filament formation is assisted by Joule heating, with local temperatures rising as high as ~500 K (*50*), whereas in our case ion transport is facilitated only by the electric field with negligible temperature increase (ΔT ~ 2 K; see SM). As a result, although very-short-range ion transport at the tip of the filament can readily occur in subnanosecond time scales, the dependence of filament properties on the local microstructure and microchemistry makes those devices unpredictable and stochastic (*2*, *9*, *49*). On the contrary, devices studied in this work are modulated by macroscopic chemical and electronic modifications, which result in distinctly controllable and deterministic performance characteristics. These benefits make their acceleration to the nanosecond regime highly relevant.

The ability to rapidly shuttle protons within bulk PSG at extreme speeds shifts the bottleneck of proton transport to the interfaces. Although the hydrogen uptake rate of Pd is high (*51*) and the PdH_x/PSG interfacial reaction is likely also efficient, the same cannot be said for proton insertion into polycrystalline WO_3 (*52*), which shows up as a PSG/H_yWO_3 interfacial charge-transfer resistance. Because protons do not have high diffusivity in H_yWO_3 (*53*), it is likely that the inserted species cannot quickly vacate their sites near the interface, thus further reducing the insertion rate in a self-limiting fashion. These factors result in an excess of highly energetic protons at the WO_3/PSG interface and may lead to H_2 gas formation and buildup at the interface (*54*).

The electrical signature and morphological consequences of long–pulse-time stressing of the devices are captured in Fig. 4. In Fig. 4A, the device previously characterized in Fig. 2 was tested under the same V_{pulse} but for increasing t_{pulse}. Above 40 ns a cascading effect was apparent, in which the conductance change increased with each pulse, ultimately causing device failure for 90-ns pulses. A scanning electron microscopy (SEM) image of the device after this experiment is shown in Fig. 4B, whereas Fig. 4C shows an image of another device stopped at earlier stages (corresponding to the 40-ns pulse regime shown in Fig. 4A). These images show damage features that we suggest were consistent with H_2 gas evolution at the PSG/H_yWO_3 interface, nanobubble formation, and stress buildup.

Most importantly, as shown in Fig. 2C, no degradation was observed under high-speed (5 ns) operation. To provide further evidence of good endurance characteristics in the ultrafast regime, we performed electron energy-loss spectroscopy (EELS) for the SE in fresh and tested devices. Figure 4D shows that the Si $L_{2,3}$ energy-loss spectra were similar in active and inactive regions across the PSG layer, which indicates that there was no stoichiometry change or that the change was too subtle to be detected. Considering that the widest electrochemical stability window for room temperature protonic electrolytes was previously ~3.35 V (*55*), we attribute our devices' ability to operate stably and reversibly at 10 V/–8.5 V to the operation at nanosecond transients, where the PSG does not have to conform to the thermodynamic stability requirement for electrolytes (established after quasi-static wait times). Furthermore, these short time scales may simply be too fast for oxygen motion (slower than that of protons), which would otherwise cause degradation of material properties. We believe these are two key physical dynamics underlying the no-damage "protonic breakdown," a transient phenomenon that we took advantage of in this novel ultrafast long-range transport regime.

In summary, we explored the limits of all–solid-state ionics to answer the question of how much faster an electrochemical artificial synapse can operate relative to its biological counterpart. Under extreme electric field conditions, we demonstrated ultrafast modulation of nanoscale protonic programmable

resistors with outstanding energy efficiency. Moreover, the devices exhibited many nonvolatile channel conductance states with good retention across a large and ideal dynamic range, which could be programmed reversibly, repeatedly, and symmetrically. Operation of the devices at extreme electric fields across the stack without any material degradation is the key breakthrough to enable such performance. Under these conditions, the activation energies that govern proton transport as well as charge-transfer reactions at the interface were substantially lowered, thus resulting in operation speed enhanced by six orders of magnitude at room temperature. Beyond artificial synapses, these findings open up possibilities in fields that require fast ion motion, such as microbatteries, artificial photosynthesis, and light-matter interactions.

REFERENCES AND NOTES

1. V. K. Sangwan, M. C. Hersam, *Nat. Nanotechnol.* **15**, 517–528 (2020).
2. Q. Xia, J. J. Yang, *Nat. Mater.* **18**, 309–323 (2019).
3. M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **1**, 22–29 (2018).
4. B. Roling, L. N. Patro, O. Burghaus, M. Gräf, *Eur. Phys. J. Spec. Top.* **226**, 3095–3112 (2017).
5. A. Röthel, S. Friedrich, R. Lühning, L. Heuer, *Z. Phys. Chem.* **224**, 1855–1889 (2010).
6. L. Onsager, *Science* **166**, 1359–1364 (1969).
7. M. J. Dignam, *J. Phys. Chem. Solids* **29**, 249–260 (1968).
8. T. Erdey-Grúz, M. Volmer, *Z. Phys. Chem.* **150A**, 203–213 (1930).
9. D. Ielmini, *IEEE Trans. Electron Dev.* **58**, 4309–4317 (2011).
10. S. Menzel, U. Böttger, M. Wimmer, M. Salinga, *Adv. Funct. Mater.* **25**, 6306–6325 (2015).
11. F. Palumbo *et al.*, *Adv. Funct. Mater.* **30**, 1900657 (2020).
12. W. Fedus, B. Zoph, N. Shazeer, arXiv:2101.03961 [cs.LG] (2021).
13. T. B. Brown *et al.*, in *Advances in Neural Information Processing Systems 33* (*NeurIPS2020*), H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, 2020), pp. 1877–1901.
14. E. Strubell, A. Ganesh, A. McCallum, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, L. Màrquez, Eds. (Association for Computational Linguistics, 2019), pp. 3645–3650.
15. I. L. Markov, *Nature* **512**, 147–154 (2014).
16. J. Choi *et al.*, in *Proceedings of the 2nd SysML Conference* (2019); https://mlsys.org/Conferences/2019/doc/2019/168.pdf.
17. X. Sun *et al.*, in *Advances in Neural Information Processing Systems 32* (*NeurIPS2019*), H. Wallach *et al.*, Eds. (Curran Associates, 2019); https://proceedings.neurips.cc/paper/2019/file/65fc9fb4897a89789352e211ca2d398f-Paper.pdf.
18. K. Steinbuch, *Kybernetik* **1**, 36–45 (1961).
19. T. Gokmen, Y. Vlasov, *Front. Neurosci.* **10**, 333 (2016).
20. T. Gokmen, M. Onen, W. Haensch, *Front. Neurosci.* **11**, 538 (2017).
21. S. Agarwal *et al.*, in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2016), pp. 929–938.
22. G. W. Burr *et al.*, *IEEE Trans. Electron Dev.* **62**, 3498–3507 (2015).
23. S. Ambrogio *et al.*, *Nature* **558**, 60–67 (2018).
24. H. Jiang *et al.*, *Sci. Rep.* **6**, 28525 (2016).
25. S. Kim *et al.*, in *2019 IEEE International Electron Devices Meeting* (IEEE, 2019), pp. 847–850.
26. Y. Li *et al.*, *Adv. Mater.* **32**, e2003984 (2020).
27. Y. van de Burgt *et al.*, *Nat. Mater.* **16**, 414–418 (2017).
28. E. J. Fuller *et al.*, *Adv. Mater.* **29**, 1604310 (2017).
29. A. Chanthbouala *et al.*, *Nat. Mater.* **11**, 860–864 (2012).
30. A. Krogh, J. A. Hertz, in *Advances in Neural Information Processing Systems 4* (*NIPS 1991*), J. Moody, S. Hanson, R. P. Lippmann, Eds. (Morgan-Kaufmann, 1991), pp. 950–957.
31. D. B. Strukov, R. S. Williams, *Appl. Phys. A* **94**, 515–519 (2009).
32. X. Yao *et al.*, *Nat. Commun.* **11**, 3134 (2020).
33. M. Onen, N. Emond, J. Li, B. Yildiz, J. A. del Alamo, *Nano Lett.* **21**, 6111–6116 (2021).
34. J. Tang *et al.*, in *2018 IEEE International Electron Devices Meeting* (IEEE, 2018).
35. A. Hjelm, C. G. Granqvist, J. M. Wills, *Phys. Rev. B* **54**, 2436–2445 (1996).
36. S. Prakash, W. E. Mustain, S. H. Park, P. A. Kohl, *J. Power Sources* **175**, 91–97 (2008).
37. M. Nogami *et al.*, *J. Electrochem. Soc.* **151**, A2095 (2004).
38. D. Bhusari, J. Li, P. J. Jayachandran, C. Moore, P. A. Kohl, *Electrochem. Solid-State Lett.* **8**, A588 (2005).
39. Y. Li, Y. T. Cheng, *Int. J. Hydrogen Energy* **21**, 281–291 (1996).
40. R. Oesten, R. A. Huggins, *Ionics* **1**, 427–437 (1995).
41. R. A. B. Devine, G. V. Herrera, *Phys. Rev. B* **63**, 233406 (2001).
42. J. Godet, A. Pasquarello, *Phys. Rev. Lett.* **97**, 155901 (2006).
43. M. E. S. Beatty, E. I. Gillette, A. T. Haley, D. V. Esposito, *ACS Appl. Energy Mater.* **3**, 12338–12350 (2020).
44. M. Kunow, A. Heuer, *J. Chem. Phys.* **124**, 214703 (2006).
45. W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, G. Ceder, *Chem. Mater.* **28**, 266–273 (2016).
46. E. H. Snow, B. E. Deal, *J. Electrochem. Soc.* **113**, 263 (1966).
47. D. Serghi, C. Pavelescu, *Thin Solid Films* **186**, L25–L28 (1990).
48. R. Vuilleumier, D. Borgis, *J. Phys. Chem. B* **102**, 4261–4264 (1998).
49. D. Ielmini, *Semicond. Sci. Technol.* **31**, 063002 (2016).
50. M. von Witzleben *et al.*, *Adv. Electron. Mater.* **3**, 1700294 (2017).
51. F. Lewis, *Platin. Met. Rev.* **4**, 132–137 (1960).
52. P. G. Dickens, S. Crouch-Baker, M. T. Weller, *Solid State Ion.* **18–19**, 89–97 (1986).
53. S. Burkhardt, M. T. Elm, B. Lani-Wayda, P. J. Klar, *Adv. Mater. Interfaces* **5**, 1701587 (2018).
54. S. Nakabayashi, R. Shinozaki, Y. Senda, H. Y. Yoshikawa, *J. Phys. Condens. Matter* **25**, 184008 (2013).
55. S. Wang *et al.*, *Adv. Mater.* **34**, 2202063 (2022).
56. M. Onen *et al.*, Data for Nanosecond Protonic Programmable Resistors for Analog Deep Learning, Zenodo (2022); doi.org/10.5281/zenodo.6506857.

ACKNOWLEDGMENTS

We thank J. Rozen, T. Todorov, P. Solomon, T. Gokmen, and M. Rasch from the IBM T. J. Watson Research Center for program support and fruitful discussions. Device fabrication took place at the facilities of MIT.nano and the MIT Electron-Beam Lithography Facility. Materials characterization took place at the facilities of MIT's Materials Research Laboratory. **Funding:** This project is funded by the MIT-IBM Watson AI Lab. N.E. would like to thank the Fonds de recherche du Québec et Nature et technologies (FRQNT) for financial support. **Author contributions:** M.O. conceived the original idea, fabricated devices, and performed experimental characterization. N.E. optimized the channel material and performed metrology studies. M.O. and N.E. designed experiments and analyzed data. B.W. designed and conducted TEM characterizations. All authors contributed to developing the theory and writing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials and can be found in Zenodo (*56*).
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abp8064
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*57*–*60*)

Submitted 26 February 2022; accepted 28 June 2022
10.1126/science.abp8064

CILIA AND FLAGELLA

In situ architecture of the ciliary base reveals the stepwise assembly of intraflagellar transport trains

Hugo van den Hoek[1,2,3]†, Nikolai Klena[4,5]†, Mareike A. Jordan[6], Gonzalo Alvarez Viar[5,6], Ricardo D. Righetto[1,2], Miroslava Schaffer[3], Philipp S. Erdmann[5], William Wan[7], Stefan Geimer[8], Jürgen M. Plitzko[3], Wolfgang Baumeister[3], Gaia Pigino[5,6]*, Virginie Hamel[4]*, Paul Guichard[4]*, Benjamin D. Engel[1,2]*

The cilium is an antenna-like organelle that performs numerous cellular functions, including motility, sensing, and signaling. The base of the cilium contains a selective barrier that regulates the entry of large intraflagellar transport (IFT) trains, which carry cargo proteins required for ciliary assembly and maintenance. However, the native architecture of the ciliary base and the process of IFT train assembly remain unresolved. In this work, we used in situ cryo–electron tomography to reveal native structures of the transition zone region and assembling IFT trains at the ciliary base in *Chlamydomonas*. We combined this direct cellular visualization with ultrastructure expansion microscopy to describe the front-to-back stepwise assembly of IFT trains: IFT-B forms the backbone, onto which bind IFT-A, dynein-1b, and finally kinesin-2 before entry into the cilium.

Cilia (and closely related flagella) are evolutionarily conserved eukaryotic organelles that extend from the cell surface and are involved in a wide variety of functions, including cell motility, fluid flow generation, sensing, and signaling (*1*, *2*). The cilium consists of an axoneme of microtubule doublets (MTDs), which extend from the microtubule triplets (MTTs) of the centriole (also called a basal body) and are sheathed in a ciliary membrane. At the base of the cilium, a specialized region known as the transition zone (TZ) gates entry and exit of both membrane-bound and soluble ciliary proteins (*3*–*5*). Among its many gating functions, the TZ is believed to regulate entry of intraflagellar transport (IFT), bidirectional traffic between the ciliary base and tip driven by kinesin and dynein motors (*6*, *7*). IFT is

[1]Biozentrum, University of Basel, 4056 Basel, Switzerland. [2]Helmholtz Pioneer Campus, Helmholtz Munich, 85764 Neuherberg, Germany. [3]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [4]Department of Molecular and Cellular Biology, Section of Biology, University of Geneva, 1211 Geneva, Switzerland. [5]Human Technopole, 20157 Milan, Italy. [6]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. [7]Department of Biochemistry and Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA. [8]Cell Biology and Electron Microscopy, University of Bayreuth, 95447 Bayreuth, Germany.
***Corresponding author. Email: gaia.pigino@fht.org (G.P.); virginie.hamel@unige.ch (V.H.); paul.guichard@unige.ch (P.G.); ben.engel@unibas.ch (B.D.E.)**
†These authors contributed equally to this work.

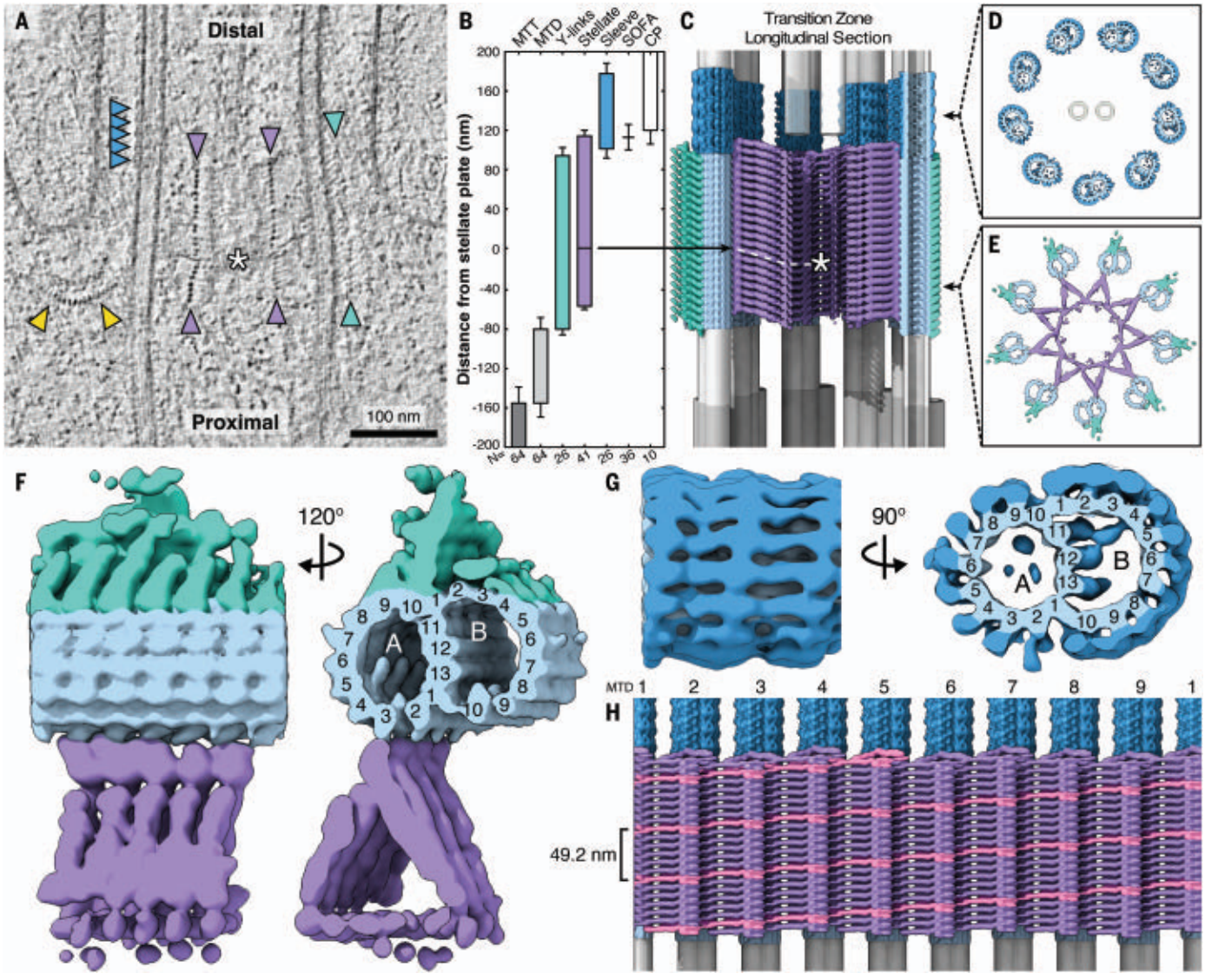


Fig. 1. Cryo-ET structure of the ciliary transition zone within native *Chlamydomonas* cells. (**A**) 2D slice through a cryo–electron tomogram, showing a TZ in longitudinal cross section. Stellate fibers, purple; Y-links, turquoise; MTD helical sleeve, dark blue; stellate plate, white asterisk (see fig. S2); and IFT, yellow (see Fig. 2). Scale bar: 100 nm. (**B**) Position along the TZ occupied by different structures. Distances are measured relative to the stellate plate (see fig. S6), defined as the "0 nm" origin point. SOFA, site of flagellar autotomy, where the cilium is cleaved (see figs. S6 and S7). CP, central pair. Error bars indicate standard deviation. (**C** to **E**) Composite model of the TZ, combining averages of the stellate (purple), Y-links (turquoise), MTD sleeve (dark blue), and associated MTD (light blue), with schematic renderings of MTTs (dark gray), MTDs (light gray), and the CP (white). (C) Longitudinal section view of the complete composite model, assembled according to the measured lengths and positions of each component, with 21 Y-link repeats, 21 stellate repeats (7 proximal of the plate, 14 distal), and 5 MTD sleeve repeats. The model shows straight MTTs and MTDs, but, as seen in (A) and quantified in (*16*), the centriole is actually a slightly convex barrel. [(D) and (E)] Cross-sectional views through the indicated regions of the composite model, showing (D) MTDs encased in the helical sleeve, with the CP in the middle, and (E) the nine-pointed stellate cylinder attached to MTDs decorated with Y-links. (**F** and **G**) Side and cross-sectional views of a single MTD attached to (F) stellate fibers and Y-links and (G) the helical sleeve. Protofilaments of the A- and B-microtubules are numbered. (**H**) Unrolled composite model, viewed from the center of the TZ, looking outward toward the inner wall of the stellate cylinder. One continuous helical density of the six-start stellate helix is marked in pink (pitch: 49.2 nm). MTDs are numbered. See also movie S1.

required for assembly and maintenance of the cilium, as well as for mediating many of its signaling functions (*8*). A pool of IFT proteins and motors resides at the ciliary base for several seconds before entry into the cilium (*9*, *10*). However, it remains to be characterized how these proteins assemble into the elaborate multi-megadalton IFT trains observed running along the ciliary axoneme (*11*). In this study, we combined in situ cryo–electron tomography (cryo-ET) (*12*) with ultrastructure expansion microscopy (U-ExM) (*13*) of the green alga *Chlamydomonas reinhardtii*, a classic model for cilia research, to reveal native TZ structures and stepwise assembly of IFT trains at the ciliary base.

Vitrified *Chlamydomonas* cells were thinned with a focused ion beam (FIB) (*14*) then imaged by cryo-ET, revealing the ciliary base inside the native cellular environment (figs. S1 and S2). Focusing on the TZ, we observed several different structures attached to the MTDs (Fig. 1, A and B), which we resolved in molecular detail by subtomogram averaging (Fig. 1, C to H,

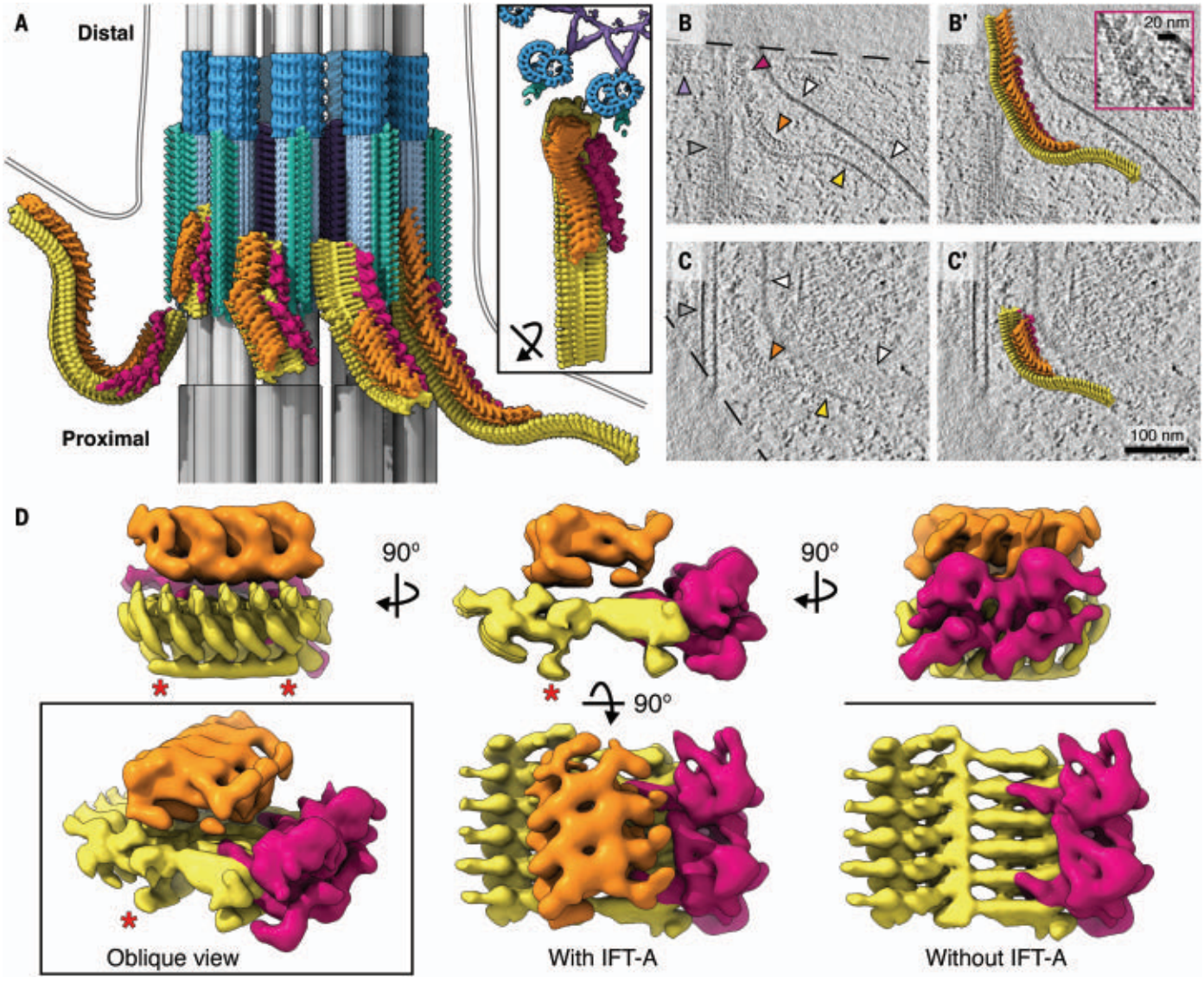


Fig. 2. Cryo-ET structures of assembling IFT trains attached to the transition zone. (**A**) Composite structures of assembling IFT trains (IFT-B, yellow; IFT-A, orange; and dynein-1b, red) engaging the ciliary TZ (colored as in Fig. 1). A schematic membrane has been added (gray double lines). Subtomogram averages have been mapped back into their positions inside a representative tomogram. (Inset) Rotated top view of the rightmost IFT train from (A). (**B** and **C**) 2D slices through tomograms, showing assembling IFT trains, and (**B′** and **C′**) overlaid 3D train structures, with their front ends attached to the TZ. (Inset) Enlarged view showing the double AAA+ adenosine triphosphatase rings of the dynein-1b motors. Arrowhead colors correspond to (A): MTT/MTD, gray, and membrane, white. Scale bars: 100 nm in (B) and (C), 20 nm in the inset. (**D**) Different views of the cytosolic IFT train composite subtomogram average. Colored as in (A). In the lower right, IFT-A was removed to show details on IFT-B. Red asterisks indicate the extra IFT-B density, which was not observed on mature trains inside cilia (see fig. S11). Assembling trains were only found attached to the TZ, not in the surrounding cytosol farther from the ciliary base (see fig. S1). See also movie S2.

and figs. S3 and S4). The proximal ~180 nm of the TZ is occupied by peripheral Y-links (Fig. 1, turquoise) and luminal stellate fibers (Fig. 1, purple, and fig. S2), which resemble a nine-pointed star in cross section (*15*). Our structure reveals that the stellates form a helical cylinder with a six-start helix and a pitch of 49.2 nm (Fig. 1H). Notably, this matches the pitch of the inner scaffold, a three-start helical cylinder in the lumen of the centriole (*16*). The nine points of the stellate star bind MTDs at protofilament A3, with a longitudinal periodicity of ~8.1 nm along each MTD (Fig. 1F and fig. S5), which matches the 8.2 nm length of one tubulin dimer.

Y-links are present in many species. They connect MTDs to the ciliary membrane and help gate transport into and out of the cilium (*17*–*20*). Our structure reveals a broad density attaching the Y-links to the MTD, spanning protofilaments A9, A10, and B1 to B4, with a longitudinal repeat of ~8.3 nm (Fig. 1F and fig. S5). However, the outer densities of the Y-links that connect to the ciliary membrane were not resolved, likely owing to flexibility of these filamentous structures. We observed that the Y-links extend along the MTDs proximal of the ciliary membrane (Fig. 1A). This suggests that connection to the membrane is not a prerequisite for MTD Y-link decoration.

Immediately distal to the stellates and Y-links, we found a distinct helical density completely decorating the MTDs (Fig. 1, dark blue). This helical "sleeve" spans 76 ± 6 nm along the MTDs, with a periodicity of ~16.4 nm (fig. S5). Cryo-ET and U-ExM immediately after ciliary

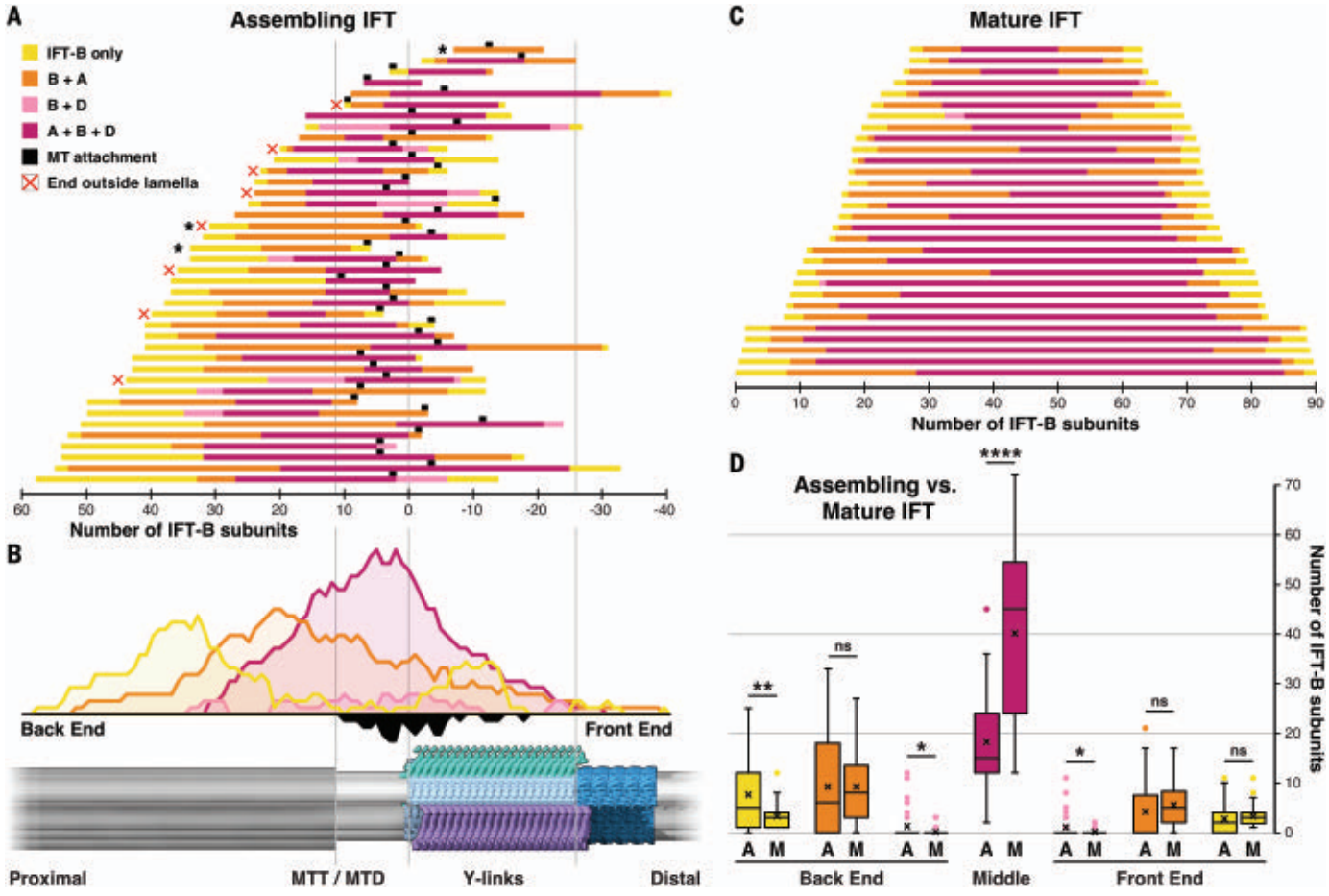


Fig. 3. IFT trains undergo stepwise assembly from front to back. (**A** and **B**) Assembly states of IFT trains relative to structures of the TZ [shown at the bottom of (B)]. Distances are measured by the number of IFT-B subunits relative to the onset of the Y-links (defined as the "0 IFT-B" origin point). (A) Position and train assembly state, measured for each IFT-B subunit of each assembling IFT train. Assembly states [color code in (A) applies to all panels] were determined on the basis of occupancy of IFT-A and/or dynein-1b on the IFT-B backbone. Black squares indicate train attachment points, that is, the most proximal IFT-B subunit that is close enough to the MTD (<20 nm) to be considered bound to the microtubule. The three trains marked with black asterisks lack a complete "middle" (no dynein is bound) and thus were omitted from analysis in (D) because "back-end B + A" could not be split from "front-end B + A." Of the 70 trains in our dataset, 30 had front ends cropped by the FIB milling and therefore were not quantified, as their starting point relative to the TZ could not be determined. Eight more trains had cropped back ends (indicated with a red "x") but were included in the analysis. (B) Cumulative plot of all assembling IFT trains in (A), with positions in relation to the TZ (shown at bottom). Colored curves above the line show the summed abundance of each IFT assembly state. The black curve below the line shows the distribution of MTD attachment points, summed from the black squares in (A). (**C**) Distribution of IFT train lengths and completeness of assembly [colored as in (A)] for mature anterograde trains found in the cilium (*11*). (**D**) Comparison of the abundance of assembly states at the front and back ends of assembling ("A") versus mature ("M") trains [i.e., trains in (A) versus trains in (C)]. Box, median and 25 to 75% percentiles; X, mean; whiskers, 1.5× interquartile range; points, outliers. Statistical significance assessed by unpaired *t* test (****$P < 0.0001$; **$P = 0.002$; *$P = 0.03$; ns, not significant, $P > 0.05$). Because the backs of some assembling trains in this analysis were clipped by the FIB milling ["x" in (A)], the "IFT-B only" region of assembling trains should be slightly more pronounced than the quantification shown here.

shedding (figs. S6 and S7) revealed that the MTD sleeve overlaps with the ciliary cleavage site, also known as the site of flagellar autotomy (SOFA) (*21*) (Fig. 1B). Consistent with this finding, the MTD sleeve was absent from the severed ends of MTDs both on the TZ (fig. S6, C and D) and isolated cilia (*22*). We therefore hypothesize that this structure might help regulate axoneme severing and is lost during the process. Because the sleeve density covers every MTD protofilament, it should sterically hinder the attachment of molecular motors and thus may also play a role in regulating IFT entry or exit. The sleeve does not appear to be present on the TZ of motile cilia from bovine trachea (*23*).

It has been known for decades that a pool of IFT proteins is localized near the base of the cilium (*24*, *25*), but the structural organization of this basal pool has remained a mystery. In our tomograms, we observed filamentous strings of particles with one end attached to the TZ and the other end splayed into the cytosol, passing between the transition fibers (Fig. 2A and fig. S8). These strings consisted of three layers, each with a different shape and periodicity (Fig. 2, B and C, and fig. S9). We performed subtomogram averaging of particles along each layer and combined them to produce a composite molecular structure (Fig. 2D and fig. S10). Comparison to a previously published cryo-ET structure of mature anterograde IFT trains within the cilium (*11*) confirmed that the cytosolic strings are indeed IFT trains. Our observed cytosolic train structure was similar to the anterograde train structure (fig. S11), enabling us to assign densities to IFT-B complexes, IFT-A complexes, and dynein-1b motors, which have longitudinal periodicities of ~6, ~11, and ~18 nm, respectively (fig. S12). As was shown for anterograde IFT, we observed

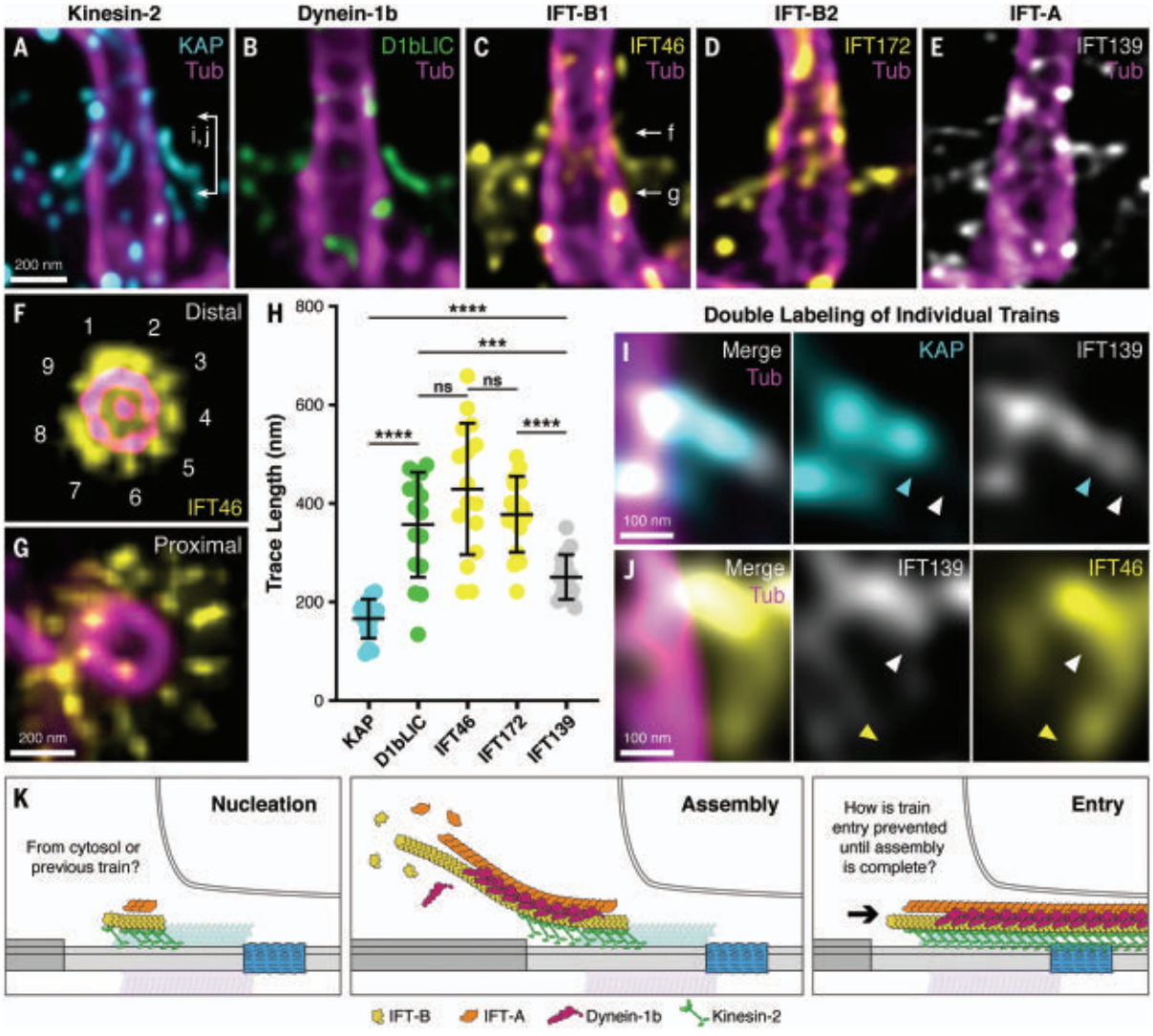


Fig. 4. Assembly of anterograde IFT trains examined by U-ExM. (**A** to **E**) Longitudinal views of assembling IFT trains at the ciliary base, visualized by antibody staining and U-ExM. (A) KAP-GFP (cyan). Connected arrows mark the approximate region shown in (I) and (J). (B) D1bLIC-GFP (green). (C) IFT46-YFP (yellow). White arrows indicate approximate positions of top views in (F) and (G). (D) IFT72 (yellow). (E) IFT139 (gray). Tubulin staining is magenta. (**F** and **G**) Top views displaying ninefold occupancy of IFT trains at (F) the distal TZ (confirmed by the presence of the central pair) and (G) the proximal TZ (see fig. S18). (**H**) Measured lengths of 3D traces for each IFT component on the assembling trains (see fig. S15E). Statistical significance assessed by unpaired *t* test (****$P < 0.0001$; ***$P = 0.0009$; ns, not significant, $P > 0.05$). U-ExM length trends are consistent with cryo-ET (Fig. 3), except that D1bLIC traces were longer than IFT139. One possible explanation could be that IFT139 is added to trains after the rest of IFT-A, but these short traces might also be influenced by a weaker anti-IFT139 antibody (see methods). (**I** and **J**) Dual staining of IFT trains at the TZ for (I) KAP + IFT139 and (J) IFT46 + IFT139. Arrowheads indicate the back ends of each IFT component on the train (see fig. S17). (**K**) Consensus model of IFT train assembly from cryo-ET and U-ExM measurements. New IFT-B subunits are added to the back of the train, forming a scaffold that is loaded with IFT-A, dynein-1b, and finally kinesin-2 close to the TZ. Z-projection thickness: 0.24 μm in (A) to (F), (I) and (J), 0.36 μm in (G). Scale bars: 200 nm in (A) to (G), 100 nm in (I) and (J).

that these cytosolic trains were attached to the MTD B-tubule (fig. S13D) (*26*) and that dynein-1b was loaded onto the trains as a cargo in an autoinhibited state (*11*, *27*). Thus, much of the IFT pool at the ciliary base consists of TZ-tethered anterograde IFT trains that are undergoing assembly before entry into the cilium. We did not see retrograde trains at the ciliary base, which implies that they rapidly disassemble, perhaps even before exiting the TZ.

We observed two notable differences between assembling and mature IFT trains. First, assembling trains are flexible and display regions of high curvature (Fig. 2, A to C, and fig. S9), whereas trains in the cilium have an extended straight conformation (fig. S12) (*11*), which is likely maintained by interactions with cargos and the ciliary membrane. Second, assembling IFT-B has a prominent extra density on the side opposite IFT-A (Fig. 2D and fig. S11; red asterisks). In regions where IFT trains are bound to the TZ, this extra density is positioned adjacent to the MTD. The density is not part of the MTD structure itself, because it is also present on IFT-B in the sections of trains that dangle into the cytosol (fig. S13).

In situ cryo-ET combines native molecular structures with precise cellular localization, enabling us to plot the spatial relationship between assembling IFT trains and the TZ (Fig. 3, A and B). The front segments of IFT trains attach to MTDs (but never to the MTTs of the centriole) and then continue in the distal direction along these microtubule tracks, while the back regions curve away from the TZ into the cytosol. The front ends of IFT trains

were frequently observed at the Y-links, but only very rarely further distal at the MTD sleeve (fig. S9). By mapping the IFT-B, IFT-A, and dynein-1b structures along the IFT trains (as shown in Fig. 2A), we visualized the spatially defined order of train assembly. Anterograde trains always contain IFT-B, which forms a backbone scaffold upon which other components are attached. This train architecture is consistent with previous findings that IFT-B recruits IFT-A into cilia (*28*, *29*). We plotted the train assembly state relative to each IFT-B subunit, and although we observed variability between individual trains (Fig. 3A), the cumulative plot shows a clear distribution of IFT train components (Fig. 3B). The back segments of trains were in a less assembled state, with IFT-B extending the farthest, followed by IFT-A and then dynein-1b (we did note some assembling trains where dynein-1b extended farther than IFT-A). No gaps were observed in the regions occupied by IFT-A and dynein-1b, suggesting that these components linearly oligomerize from front to back on the IFT-B scaffold.

Next, we compared the assembling cytosolic trains to mature anterograde trains within the cilium (Fig. 3, C and D). Mature trains exhibited a continuum of lengths from 223 to 558 nm, comprising 36 to 90 IFT-B subunits with an average of 62 ± 16 IFT-B. In comparison, the longest assembling train we observed contained 88 IFT-B subunits, and the average length was 44 ± 17 IFT-B. Thus, just one IFT train is assembled at a time per MTD. Cytosolic and ciliary IFT trains have similarly incomplete front ends, whereas the incomplete segment at the back of cytosolic trains is longer, as these regions are still undergoing assembly (Fig. 3D). Perhaps as IFT trains enter the cilium, part of the incompletely assembled back end is left behind at the TZ to nucleate the front of the next train (Fig. 4K). Because assembly of IFT-A and dynein-1b proceeds front to back, the front end of the new train would remain incomplete, as we see in mature anterograde trains inside the cilium (Fig. 3C).

The kinesin-2 anterograde IFT motor is difficult to identify by cryo-ET owing to its relatively small size and flexible structure. We thus turned to U-ExM to examine the occupancy of kinesin-2 on assembling trains. We imaged components of kinesin-2, dynein-1b, IFT-B (subcomplexes B1 and B2), and IFT-A (Fig. 4, A to G, and fig. S14). Consistent with our cryo-ET data, all of these proteins formed filamentous strings at the ciliary base, with their distal ends in close proximity to the TZ (marked by centrin staining and the lack of glutamylated tubulin) (figs. S15 and S16) and their proximal ends curving into the cytoplasm. Using semiautomated tracing, we plotted the three-dimensional (3D) trajectories of each protein (fig. S15E). IFT-B traces were longer than traces of IFT-A and dynein-1b (Fig. 4H). Kinesin-2 traces were the shortest and were localized at the front of IFT trains close to the TZ (fig. S16). Double-labeling of individual IFT trains confirmed that IFT-A extends farther than kinesin-2 toward the back of the train, and IFT-B extends farther than IFT-A (Fig. 4, I and J, and fig. S17).

We observed ninefold occupancy of IFT around the majority of TZs (Fig. 4, F and G, and fig. S18). Measured by U-ExM, cryo-ExM, and cryo-ET, 93 to 96% of MTDs were bound by an assembling IFT train. Given that trains enter the cilium with a frequency of ~1 train/s (*28*, *30*), we predict that trains linger at the ciliary base for an average of 9 s during assembly. Consistent with this hypothesis, photobleaching analysis of the *Chlamydomonas* basal pool showed that several green fluorescent protein (GFP)–tagged IFT proteins require ~9 s to recover peak fluorescence (*9*). Regardless, ciliary entry appears to be stochastic rather than sequential, with an avalanche-like relationship between lag time and IFT train size (*31*).

Our combined cryo-ET and U-ExM measurements support a spatially defined sequence of IFT train assembly: IFT-B first forms the backbone, which scaffolds the front-to-back attachment of IFT-A, dynein-1b, and finally kinesin-2 close to the TZ (Fig. 4K). A major remaining question is how IFT entry into the cilium is regulated. What prevents the front end of the train from entering the cilium as assembly proceeds toward the incomplete back of the train? Possible mechanisms may include controlled loading of kinesin-2 by factors at the TZ, the sleeve serving as a barrier to prevent entry of small trains, or the extra IFT-B–associated density acting as a molecular brake until train assembly is completed.

REFERENCES AND NOTES

1. M. V. Nachury, D. U. Mick, *Nat. Rev. Mol. Cell Biol.* **20**, 389–405 (2019).
2. D. R. Mitchell, *Cold Spring Harb. Perspect. Biol.* **9**, a028290 (2017).
3. Q. Hu *et al*., *Science* **329**, 436–439 (2010).
4. H. L. Kee *et al*., *Nat. Cell Biol.* **14**, 431–437 (2012).
5. D. K. Breslow, E. F. Koslover, F. Seydel, A. J. Spakowitz, M. V. Nachury, *J. Cell Biol.* **203**, 129–147 (2013).
6. G. Pigino, *Curr. Biol.* **31**, R530–R536 (2021).
7. K. G. Kozminski, K. A. Johnson, P. Forscher, J. L. Rosenbaum, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5519–5523 (1993).
8. H. Ishikawa, W. F. Marshall, *Nat. Rev. Mol. Cell Biol.* **12**, 222–234 (2011).
9. J. L. Wingfield *et al*., *eLife* **6**, e26609 (2017).
10. J. V. K. Hibbard, N. Vazquez, R. Satija, J. B. Wallingford, *Mol. Biol. Cell* **32**, 1171–1180 (2021).
11. M. A. Jordan, D. R. Diener, L. Stepanek, G. Pigino, *Nat. Cell Biol.* **20**, 1250–1255 (2018).
12. S. Asano, B. D. Engel, W. Baumeister, *J. Mol. Biol.* **428**, 332–343 (2016).
13. D. Gambarotto *et al*., *Nat. Methods* **16**, 71–74 (2019).
14. M. Schaffer *et al*., *J. Struct. Biol.* **197**, 73–82 (2017).
15. S. Geimer, M. Melkonian, *J. Cell Sci.* **117**, 2663–2674 (2004).
16. M. Le Guennec *et al*., *Sci. Adv.* **6**, eaaz4137 (2020).
17. B. Craige *et al*., *J. Cell Biol.* **190**, 927–940 (2010).
18. C. L. Williams *et al*., *J. Cell Biol.* **192**, 1023–1041 (2011).
19. F. R. Garcia-Gonzalo, J. F. Reiter, *Cold Spring Harb. Perspect. Biol.* **9**, a028134 (2017).
20. A. R. G. De-Castro *et al*., *J. Cell Biol.* **221**, e202010178 (2022).
21. L. M. Quarmby, in *The Chlamydomonas Sourcebook* (*Second Edition*), *Volume 3*, E. H. Harris, D. B. Stern, G. B. Witman, Eds. (Academic Press, 2009), pp. 43–69.
22. K. H. Bui, T. Yagi, R. Yamamoto, R. Kamiya, T. Ishikawa, *J. Cell Biol.* **198**, 913–925 (2012).
23. G. A. Greenan, R. D. Vale, D. A. Agard, *J. Cell Biol.* **219**, e201907060 (2020).
24. D. G. Cole *et al*., *J. Cell Biol.* **141**, 993–1008 (1998).
25. Y. Hou *et al*., *J. Cell Biol.* **176**, 653–665 (2007).
26. L. Stepanek, G. Pigino, *Science* **352**, 721–724 (2016).
27. K. Toropova *et al*., *Nat. Struct. Mol. Biol.* **26**, 823–829 (2019).
28. C. Iomini, V. Babaev-Khaimov, M. Sassaroli, G. Piperno, *J. Cell Biol.* **153**, 13–24 (2001).
29. J. M. Brown, D. A. Cochran, B. Craige, T. Kubo, G. B. Witman, *Curr. Biol.* **25**, 1583–1593 (2015).
30. B. D. Engel, W. B. Ludington, W. F. Marshall, *J. Cell Biol.* **187**, 81–89 (2009).
31. W. B. Ludington, K. A. Wemmer, K. F. Lechtreck, G. B. Witman, W. F. Marshall, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3925–3930 (2013).

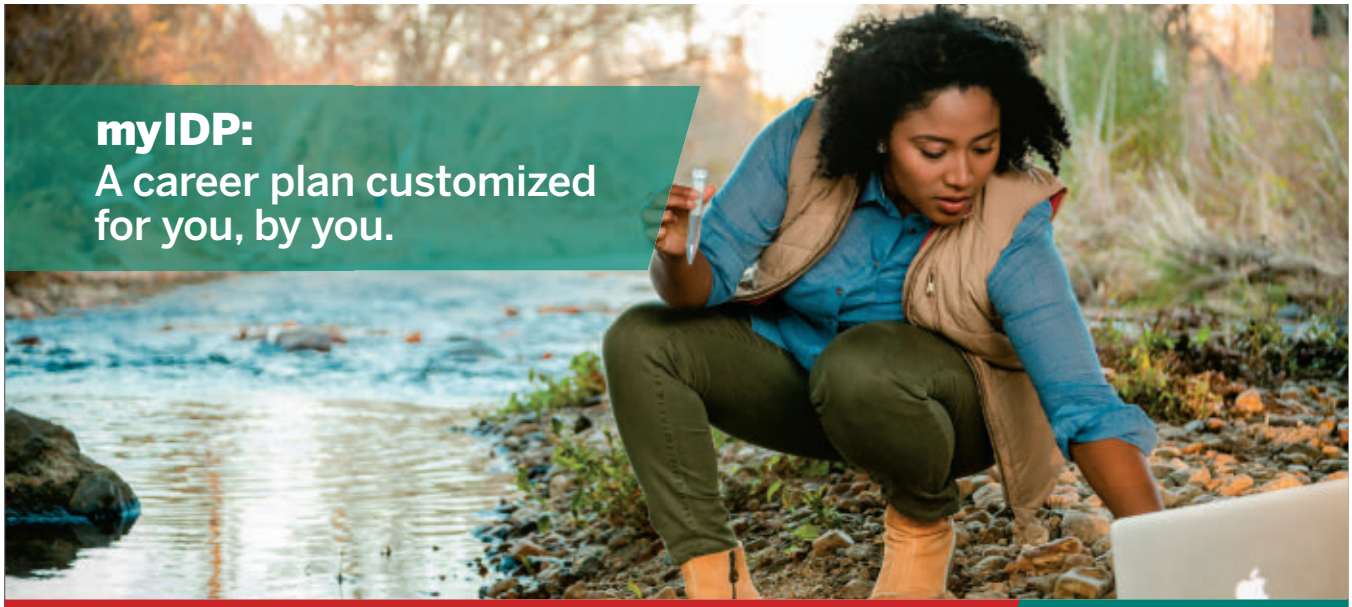
ACKNOWLEDGMENTS

We thank T. Ishikawa, A. Noga, D. Diener, and K. Lechtreck for helpful discussions that influenced the manuscript. G.P. and G.A.V. thank T. Furstenhaupt from the MPI-DBG EM facility for instrument support. Calculations were performed at the Max Planck Institute for Biochemistry computing cluster in Martinsried, Germany, and at the sciCORE (https://scicore.unibas.ch/) scientific computing center at the University of Basel. **Funding:** This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant 819826) to G.P., the Swiss National Science Foundation (SNSF) Grant PP00P3_187198 to P.G., the ERC ACCENT Starting Grant 715289 to P.G., a European Molecular Biology Organization (EMBO) Fellowship (grant ALTF 537–2021) to N.K., and an Alexander von Humboldt Foundation Fellowship to R.D.R. (R.D.R. also acknowledges a nonstipendiary fellowship from EMBO). Additional funding for personnel and instrumentation was provided by Helmholtz Munich, the Max Planck Society, and the Biozentrum of the University of Basel. **Author contributions:** M.S., P.S.E., and B.D.E. acquired cryo-ET data of FIB-milled *C. reinhardtii* cells. G.A.V. prepared samples and acquired cryo-ET data of isolated basal bodies. M.A.J. prepared samples and acquired cryo-ET data of intact cilia. H.v.d.H. analyzed all the cryo-ET data, with help from M.A.J., R.D.R., and W.W. U-ExM data acquisition and analysis was performed by N.K., with guidance from P.G. and V.H. Conventional electron microscopy was performed by S.G. Additional mentorship was provided by W.B. and J.M.P., along with access to instrumentation. G.P., V.H., P.G., and B.D.E. initiated and supervised the study. H.v.d.H., N.K., and B.D.E. wrote the paper, with input from all authors. **Competing interests:** W.B. is on the life science advisory board of Thermo Fisher. **Data and materials availability:** Cellular tomograms, subtomogram averages, and composite maps have been deposited at the Electron Microscopy Data Bank (EMD-15252 to EMD-15263, see table S2). Raw cryo-ET data are available from the Electron Microscopy Public Image Archive (EMPIAR-11078). All data needed to evaluate the conclusions of the study are present in the paper and the supplementary materials. Correspondence and requests for materials should be addressed to B.D.E. (ben.engel@unibas.ch). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6704
Materials and Methods
Figs. S1 to S18
Tables S1 and S2
References (*32*–*64*)
MDAR Reproducibility Checklist
Movies S1 and S2

Submitted 3 October 2021; accepted 27 June 2022
10.1126/science.abm6704



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
UMASS University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

Science Careers online @sciencecareers.org

WORKING LIFE

By **Ahmed Elbassiouny**

Unequal undertakings

When I started graduate school, I knew it was not going to be an easy ride—not only because of the work, but also financially. My graduate stipend is thousands of dollars below the federal poverty line. This is perhaps adequate when graduate students have family support to fall back on, or at least minimal expenses. But for me, and for many others, this couldn't be further from the truth. I left behind my family, friends, and support system in Egypt when I was 17 to pursue higher education in Canada and chase the dream of working alongside world-renowned scientists. I also have a chronic disease that knows no break. To stay alive, I rely on daily medications and supplies that are not fully covered by the national health care system. My graduate stipend is barely enough to pay for basic living costs, and my medical expenses are nearly impossible to afford.

To make ends meet, I have taken on extra hours working as a teaching assistant. I'm lucky this is even an option for me; in many departments, graduate students are not allowed to hold extra teaching appointments as it "distracts" from thesis research. The prevailing assumption seems to be that all students can and should dedicate every minute to their thesis, and failure to do so is a sign of lack of devotion or drive.

But my teaching has nothing to do with a lack of passion for my field of study. I love my research. Sometimes, when I go home after setting up an exciting experiment, I find myself counting the hours until I can return to the lab and see the results the next morning. What distracts me from my science is not my teaching per se, but my financial worries.

"What distracts me from my science is not my teaching per se, but my financial worries."

Even though my department allows me to take on extra teaching, my peers and advisers often judge me for doing it. In one particularly memorable instance, a junior lab mate asked about my plans for the summer, and I mentioned I was going to be teaching in addition to presenting at a conference, conducting experiments, and doing everything else involved in pursuing a Ph.D. "Again?!" he exclaimed. "You are teaching way too much! You should spend the summer focusing on your research and increasing your productivity." He had no idea what I was juggling.

I don't tell my colleagues about my health condition because I don't want it to define me or my scientific capabilities. I thought working hard to hide my struggles and meet expectations was the best way forward. At one point, I was teaching three times the expected number of hours while also spending late nights and weekends on my thesis work. I put on a brave face, but in reality I was scared, lonely, and exhausted.

Eventually my unhealthy hamster wheel got to me. My health suffered significantly and I developed complications. I spent the latter half of 2019 and the first years of the pandemic running from one specialist to another, at one point requiring biweekly invasive eye procedures that left me in debilitating pain. Still, I continued to work on my thesis, returning to the lab as soon as my blurry vision returned to normal. I was trying to devote everything to my academic journey, like my peers who have easier circumstances—despite my health condition, greater expenses, and lack of family support.

As I approach the final stretch of my Ph.D., I look forward to leaving academia for a job where my efforts are appreciated and my well-being respected. I am grateful for the scientific training I have received, but the cost has been too high. To truly foster inclusivity in academia, we must understand that academics have outside commitments including taking care of children, parents, and our own health. We should be helped through those challenges—for example, with less humiliating pay and reasonable work expectations—instead of being judged for being insufficiently dedicated. Inclusion means ensuring everyone feels comfortable, valued, and supported, whatever our circumstances. ■

Ahmed Elbassiouny is a Ph.D. student at the University of Toronto.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

NEWS FROM Science | AAAS

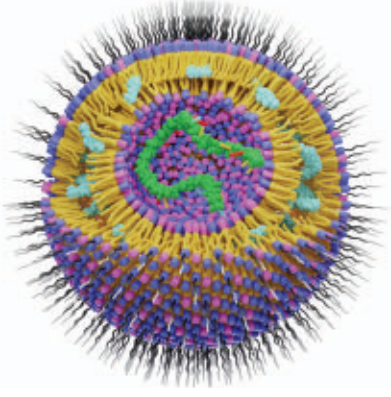
Subscribe to **News from Science** for unlimited access to authoritative, up-to-the-minute news on research and science policy.

bit.ly/NewsFromScience

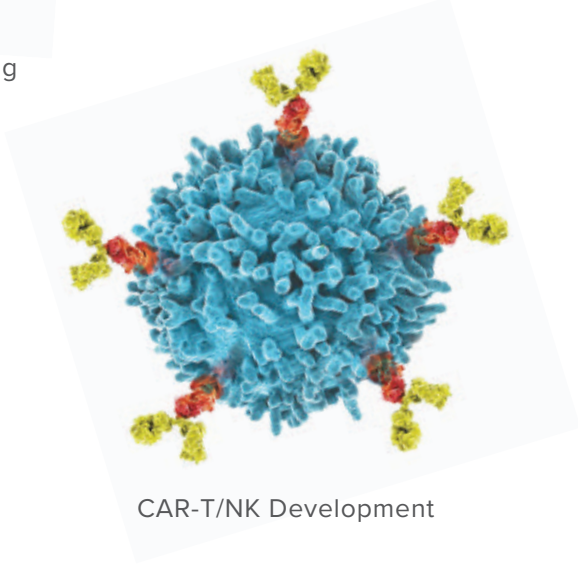
Expanding Immunotherapy & Vaccine Research

Harnessing Antibodies for Oncology & Immunology

Antibody Engineering



mRNA-LNP Formulation



CAR-T/NK Development

Antibody Services

- Monoclonal Ab Development
- Hybridoma Sequencing
- Human Antibody Screening
- Antibody Humanization
- Bispecific Antibody Engineering
- Stable Cell Line Generation

Immune Cell Services

- Lentivirus Production GMP
- CAR-T/NK Cell Generation
- CAR-T/NK Animal Model
- Immune Cell Engineering
- T/B/NK/Dendritic cell, Macrophages
- TCR knockout by CRISPR

mRNA-LNP Services

- mRNA Design and Synthesis (IVT)
- mRNA Capping & Purification
- mRNA-LNP Encapsulation
- mRNA-LNP Scale Up and QC
- mRNA-LNP Validation In Vivo
- Ionizable Lipid Licensing

All products are for research use only

Discover more | www.promab.com

ProMab Biotechnologies, Inc.
2600 Hilltop Dr, Building B, Suite C320, Richmond, CA 94806
1.866.339.0871 | info@promab.com

SCAN ME

Pandemic preparedness in a changing world: Fostering global collaboration to strengthen public health

A sponsored supplement to *Science*

Produced by the *Science*/AAAS Custom Publishing Office

Sponsored by

A sponsored supplement to *Science*

Pandemic preparedness in a changing world:

Fostering global collaboration to strengthen public health and respond to viral threats

Sponsored by

Abbott

Produced by the *Science*/AAAS Custom Publishing Office

Science | AAAS

science.org/journal/sciimmunol

READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology

Table of Contents

This supplement was produced by the *Science*/AAAS Custom Publishing Office and sponsored by Abbott.

The original content articles from the sponsor and their partners that appear in this supplement have not been peer-reviewed nor have they been assessed by *Science*. Additionally, those articles marked as "News" have not been peer-reviewed.

Editor: Sean Sanders, Ph.D.
Proofreader/Copyeditor: Bob French
Designer: JD Huntsinger

Chris Hoag
Sales Manager
choag@aaas.org
+1 (202) 590-8176

© 2022 by The American Association for the Advancement of Science. All rights reserved.
29 July 2022

Introductions

2 **Be prepared**
Sean Sanders, Ph.D.
Science/AAAS

3 **Fighting future viruses, together**
John R. Hackett Jr., Ph.D.
Abbott

Section 1: Research and News Articles

5 **A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa**
Eduan Wilkinson, Marta Giovanetti, Houriiyah Tegally *et al.*

14 **Monkeypox is a new global threat. African scientists know what the world is up against**
Jon Cohen

16 **The puzzle of the COVID-19 pandemic in Africa**
Justin M. Maeda and John N. Nkengasong

18 **SARS-like viruses may jump from animals to people hundreds of thousands of times a year**
Kai Kupferschmidt

19 **The Global Virome Project**
Dennis Carroll, Peter Daszak, Nathan D. Wolfe *et al.*

21 **COVID-19 testing: One size does not fit all**
Michael J. Mina and Kristian G. Andersen

23 **Antimicrobial resistance: A global challenge**
Simon J. Howard, Sarah Hopwood, Sally C. Davies

25 **Modeling infectious disease dynamics in the complex landscape of global health**
Hans Heesterbeek, Roy M. Anderson, Viggo Andreasen *et al.*

35 **Computer scan uncovers 100,000 new viruses**
Elizabeth Pennisi

36 **A highly virulent variant of HIV-1 circulating in the Netherlands**
Chris Wymant, Daniela Bezemer, François Blanquart *et al.*

Section 2: Original Content from Sponsor and Partners

43 **Pandemic defense: The key is global collaboration**
Gavin Cloherty

45 **Helping to prevent the next pandemic: The Abbott Pandemic Defense Coalition**

47 **Protecting public health through viral molecular surveillance**
Mary A. Rodgers

53 **Virus discovery for diagnostic preparedness**
Michael G. Berg

60 **The key role of field epidemiology in preventing the next pandemic: The partnership between Abbott and Training Programs in Epidemiology and Interventions Network (TEPHINET)**
Francisco Averhoff, Lara Pereira, Susan Gawel

63 **The case for networking the networks**
Gavin Cloherty, Nicaise Ndembi, Ian Lipkin, Sandra Chavez

67 **Building genomic sequencing capacity in Africa to respond to the SARS-CoV-2 pandemic**
Tulio de Oliveira, Eduan Wilkinson, Cheryl Baxter *et al.*

Be prepared

The founder of the global Scouting movement, Robert Baden-Powell, famously cautioned his young charges to always "be prepared." While excellent advice for those trekking into the wilderness, is it applicable in a broader sense to a world filled with complexities and webs of interdependency? I would argue it is, particularly when it comes to protecting the health of the global population by being properly prepared for the next pandemic.

In September 2021, the White House released a short but important document outlining the administration's goals for future pandemic preparedness.[1] It will no doubt require iterative revisions in the future to keep up with a changing world, but it does provide a good, immediate starting point for the United States—and other countries—to ready ourselves for the inevitable next virus outbreak.

The key goals of this report are: (1) transform our medical defenses, (2) ensure situational awareness, (3) strengthen public health systems, (4) build core capabilities, and (5) manage the mission. It is the second goal that I believe provides the opportunity to have the greatest impact and immediate benefit for pandemic preparedness.

This situational awareness goal includes the creation of rapid detection systems for identifying viruses with pandemic potential and the real-time monitoring of current outbreaks that could transform into pandemics. Consider how many lives might have been saved, not to mention unnecessary suffering of patients and their families, if an effective, transparent early-warning system was in place when the SARS-CoV-2 virus began its rampage? We have gotten better at real-time monitoring of COVID variants, but there is much room for improvement.

To state the obvious, creating, maintaining, and managing such systems will not be an easy task. There will need to be increased cooperation between all countries, true transparency, and clear accountability. Industry, academia, and government agencies will need to build strong ties and greater trust than ever before. The financial commitment will be considerable. But when compared with the cost to society in terms of lost lives and battered economies, it will be a relatively small price to pay.

Now is the time to act, while the impacts of the COVID pandemic are still fresh in our minds and the political will is still strong. Let's follow the lead of those young Scouts and commit to earning our "Be Prepared" badge.

Sean Sanders, Ph.D.
Science/AAAS

[1]https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf.

Fighting future viruses, together

Fighting pandemics is a team sport.

We must have a strong offense in play at all times to anticipate and prepare for what's coming. And we need a fast, effective defense to stop and neutralize viruses before they spread uncontrollably.

The world is now living in an era where epidemic and pandemic threats have become the new normal. If there is a silver lining in the COVID-19 pandemic, it's what it taught us: Global collaboration is the key to stopping future pandemics.

At Abbott, we are focused on helping people around the world live their healthiest lives, and doing so means preventing another pandemic from ever happening again. We are proud of our work in the fight against COVID-19 as the leader in testing solutions across lab, point-of-care, and rapid settings. But we are still playing defense against COVID-19 as it continues to evolve and as new viral threats emerge.

We have to stay one step ahead, which requires strong collaboration with public and private partners everywhere. We have a team of virus hunters with decades of experience that works day in and day out looking for new threats.

Abbott has nearly 30 years of foundational experience hunting and fighting viruses, and an even longer history of developing diagnostics tests. From inventing the first U.S. Food and Drug Administration–approved HIV test in 1985 to establishing our first Global Surveillance Program in 1994, the company's history has led us to this moment, and we are proud to continue building on our previous contributions.

Our virus hunters are now part of the Abbott Pandemic Defense Coalition, a first-of-its-kind global partnership launched last year with the goal of actively working to prevent the next pandemic. Today, coalition partners are on the ground in 11 countries across five continents, actively collecting, logging, and monitoring diverse pathogens to keep pace with new and changing viruses.

A global pandemic such as COVID-19 is bigger than any of us—it requires that we come together as public health experts, governments, and private organizations across the world. We need to share information so we can track how these threats are spreading and prevent them from becoming pandemics in the first place.

Now is the time to unite and act quickly and effectively as a global virus-fighting team. We know it's not a matter of if another pandemic will happen, but when.

Together, working side by side, we can be ready.

John R. Hackett Jr., Ph.D.
Divisional Vice President of Applied Research and Technology
Abbott

Section One: Research and News Articles

Relevant and seminal articles from the *Science* family of journals

RESEARCH ARTICLES

CORONAVIRUS

A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa

Eduan Wilkinson[1,2]†, Marta Giovanetti[3,4]†, Houriiyah Tegally[1]†, James E. San[1]†, Richard Lessells[1], Diego Cuadros[5], Darren P. Martin[6,7], David A. Rasmussen[8,9], Abdel-Rahman N. Zekri[10], Abdoul K. Sangare[11], Abdoul-Salam Ouedraogo[12], Abdul K. Sesay[13], Abechi Priscilla[14], Adedotun-Sulaiman Kemi[14], Adewunmi M. Olubusuyi[15], Adeyemi O. O. Oluwapelumi[16], Adnène Hammami[17], Adrienne A. Amuri[18,19], Ahmad Sayed[20], Ahmed E. O. Ouma[21], Aida Elargoubi[22,23], Nnennaya A. Ajayi[24], Ajogbasile F. Victoria[14], Akano Kazeem[14], Akpede George[25], Alexander J. Trotter[26], Ali A. Yahaya[27], Alpha K. Keita[28,29], Amadou Diallo[30], Amadou Kone[31], Amal Souissi[32], Amel Chtourou[17], Ana V. Gutierrez[26], Andrew J. Page[26], Anika Vinze[33], Arash Iranzadeh[6,7], Arnold Lambisia[34], Arshad Ismail[35], Audu Rosemary[36], Augustina Sylverken[37], Ayoade Femi[14], Azeddine Ibrahimi[38], Baba Marycelin[39], Bamidele S. Oderinde[39], Bankole Bolajoko[14], Beatrice Dhaala[40], Belinda L. Herring[27], Berthe-Marie Njanpop-Lafourcade[27], Bronwyn Kleinhans[41], Bronwyn McInnis[10], Bryan Tegomoh[42], Cara Brook[43,44], Catherine B. Pratt[45], Cathrine Scheepers[35,46], Chantal G. Akoua-Koffi[47], Charles N. Agoti[34,48], Christophe Peyrefitte[30], Claudia Daubenberger[49], Collins M. Morang'a[50], D. James Nokes[34,51], Daniel G. Amoako[35], Daniel L. Bugembe[40], Danny Park[33], David Baker[26], Deelan Doolabh[7], Deogratius Ssemwanga[40,52], Derek Tshiabuila[1], Diarra Bassirou[30], Dominic S. Y. Amuzu[50], Dominique Goedhals[53], Donwilliams O. Omuoyo[34], Dorcas Maruapula[54], Ebenezer Foster-Nyarko[26], Eddy K. Lusamaki[18,19], Edgar Simulundu[55], Edidah M. Ong'era[34], Edith N. Ngabana[18,19], Edwin Shumba[56], Elmostafa El Fahime[57], Emmanuel Lokilo[18], Enatha Mukantwari[58], Eromon Philomena[14], Essia Belarbi[59], Etienne Simon-Loriere[60], Etilé A. Anoh[47], Fabian Leendertz[59], Faida Ajili[61], Fakayode O. Enoch[62], Fares Wasfi[63], Fatma Abdelmoula[32,64], Fausta S. Mosha[27], Faustinos T. Takawira[65], Fawzi Derrar[66], Feriel Bouzid[32], Folarin Onikepe[14], Fowotade Adeola[67], Francisca M. Muyembe[18,19], Frank Tanser[68,69,70], Fred A. Dratibi[27], Gabriel K. Mbunsu[19], Gaetan Thilliez[26], Gemma L. Kay[26], George Githinji[34,71], Gert van Zyl[41,72], Gordon A. Awandare[50], Grit Schubert[59], Gugu P. Maphalala[73], Hafaliana C. Ranaivoson[44], Hajar Lemriss[74], Happi Anise[14], Haruka Abe[75], Hela H. Karray[17], Hellen Nansumba[76], Hesham A. Elgahzaly[77], Hlanai Gumbo[65], Ibtihel Smeti[32], Ikhlas B. Ayed[32], Ikponmwosa Odia[25], Ilhem Boutiba Ben Boubaker[78,79], Imed Gaaloul[22], Inbal Gazy[80], Innocent Mudau[7], Isaac Ssewanyana[76], Iyaloo Konstantinus[81], Jean B. Lekana-Douk[82], Jean-Claude C. Makangara[18,19], Jean-Jacques M. Tamfum[18,19], Jean-Michel Heraud[30,44], Jeffrey G. Shaffer[83], Jennifer Giandhari[1], Jingjing Li[84], Jiro Yasuda[75], Joana Q. Mends[85], Jocelyn Kiconco[52], John M. Morobe[34], John O. Gyapong[85], Johnson C. Okolie[14], John T. Kayiwa[40], Johnathan A. Edwards[68,86], Jones Gyamfi[85], Jouali Farah[87], Joweria Nakaseegu[52], Joyce M. Ngoi[50], Joyce Namulondo[52], Julia C. Andeko[82], Julius J. Lutwama[40], Justin O'Grady[26], Katherine Siddle[33], Kayode T. Adeyemi[14], Kefentse A. Tumedi[88], Khadija M. Said[34], Kim Hae-Young[89], Kwabena O. Duedu[85], Lahcen Belyamani[38], Lamia Fki-Berrajah[17], Lavanya Singh[1], Leonardo de O. Martins[26], Lynn Tyers[7], Magalutcheemee Ramuth[91], Maha Mastouri[22,23], Mahjoub Aouni[22], Mahmoud el Hefnawi[92], Maitshwarelo I. Matsheka[88], Malebogo Kebabonye[93], Mamadou Diop[30], Manel Turki[32], Marietou Paye[33], Martin M. Nyaga[94], Mathabo Mareka[95], Matoke-Muhia Damaris[96], Maureen W. Mburu[34], Maximillian Mpina[49,97,98], Mba Nwando[99], Michael Owusu[100], Michael R. Wiley[45], Mirabeau T. Youtchou[101], Mitoha O. Ayekaba[97], Mohamed Abouelhoda[102,103], Mohamed G. Seadawy[104], Mohamed K. Khalifa[20], Mooko Sekhele[95], Mouna Ouadghiri[38], Moussa M. Diagne[30], Mulenga Mwenda[105], Mushal Allam[35], My V. T. Phan[40], Nabil Abid[79,106], Nadia Touil[107], Nadine Rujeni[108,109], Najla Kharrat[32], Nalia Ismael[110], Ndongo Dia[30], Nedio Mabunda[110], Nei-yuan Hsiao[7,111], Nelson B. Silochi[97], Ngoy Nsenga[27], Nicksy Gumede[27], Nicola Mulder[112], Nnaemeka Ndodo[99], Norosoa H Razanajatovo[44], Nosamiefan Iguosadolo[14], Oguzie Judith[14], Ojide C. Kingsley[113], Okogbenin Sylvanus[25], Okokhere Peter[25], Oladiji Femi[114], Olawoye Idowu[14], Olumade Testimony[14], Omoruyi E. Chukwuma[67], Onwe E. Ogah[115], Chika K. Onwuamah[36,138], Oshomah Cyril[25], Ousmane Faye[30], Oyewale Tomori[14], Pascale Ondoa[56], Patrice Combe[116], Patrick Semanda[76], Paul E. Oluniyi[14], Paulo Arnaldo[110], Peter K. Quashie[50], Philippe Dussart[44], Phillip A. Bester[53], Placide K. Mbala[18,19], Reuben Ayivor-Djanie[85], Richard Njouom[117], Richard O. Phillips[118], Richmond Gorman[118], Robert A. Kingsley[26], Rosina A. A. Carr[85], Saâd El Kabbaj[119], Saba Gargouri[17], Saber Masmoudi[32], Safietou Sankhe[30], Salako B. Lawal[36], Samar Kassim[77], Sameh Trabelsi[120], Samar Metha[33], Sami Kammoun[121], Sanaâ Lemriss[122], Sara H. A. Agwa[77], Sébastien Calvignac-Spencer[59], Stephen F. Schaffner[33], Seydou Doumbia[31], Sheila M. Mandanda[18,19], Sherihane Aryeetey[123], Shymaa S. Ahmed[123], Siham Elhamoumi[33], Soafy Andriamandimby[44], Sobajo Tope[14], Sonia Lekana-Douki[82], Sophie Prosolek[26], Soumeya Ouangraoua[124,125], Steve A. Mundeke[18,19], Steven Rudder[26], Sumir Panji[112], Sureshnee Pillay[1], Susan Engelbrecht[41,72], Susan Nabadda[76], Sylvie Behillil[126], Sylvie L. Budiaki[95], Sylvie van der Werf[126], Tapfumanei Mashe[65], Tarik Aanniz[38], Thabo Mohale[35], Thanh Le-Viet[26], Tobias Schindler[49,97], Ugochukwu J. Anyaneji[1], Ugwu Chinedu[14], Upasana Ramphal[1,69,127], Uwanibe Jessica[14], Uwem George[14], Vagner Fonseca[1,4,128], Vincent Enouf[126], Vivianne Gorova[129,130], Wael H. Roshdy[123], William K. Ampofo[50], Wolfgang Preiser[41,72], Wonderful T. Choga[54,131], Yaw Bediako[50], Yeshnee Naidoo[1], Yvan Butera[108,132,133], Zaydah R. de Laurent[34], Amadou A. Sall[30], Ahmed Rebai[32], Anne von Gottberg[35,139], Bourema Kouriba[12], Carolyn Williamson[7,69,111], Daniel J. Bridges[105], Ihekweazu Chikwe[99], Jinal N. Bhiman[35,139], Madisa Mine[134], Matthew Cotten[40,135], Sikhulile Moyo[54,136], Simani Gaseitsiwe[54,136], Ngonda Saasa[55], Pardis C. Sabeti[33], Pontiano Kaleebu[40], Yenew K. Tebeje[21], Sofonias K. Tessema[21], Christian Happi[14], John Nkengasong[21], Tulio de Oliveira[1,2,69,137]*

The progression of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic in Africa has so far been heterogeneous, and the full impact is not yet well understood. In this study, we describe the genomic epidemiology using a dataset of 8746 genomes from 33 African countries and two overseas territories. We show that the epidemics in most countries were initiated by importations predominantly from Europe, which diminished after the early introduction of international travel restrictions. As the pandemic progressed, ongoing transmission in many countries and increasing mobility led to the emergence and spread within the continent of many variants of concern and interest, such as B.1.351, B.1.525, A.23.1, and C.1.1. Although distorted by low sampling numbers and blind spots, the findings highlight that Africa must not be left behind in the global pandemic response, otherwise it could become a source for new variants.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late 2019 in Wuhan, China (*1*, *2*). Since then, the virus has spread to all corners of the world, causing almost 150 million cases of COVID-19 and more than 3 million deaths by the end of April 2021. Throughout the pandemic, it has been noted that Africa accounts for a relatively low proportion of reported cases and deaths—by the end of April 2021, there had been ~4.5 million cases and ~120,000 deaths on the continent, corresponding to less than 4% of the global burden. However, emerging data from seroprevalence surveys and autopsy studies in some African countries suggest that the true number of infections and deaths may be severalfold higher than reported (*3*, *4*). In addition, a recent analysis has shown that in

many African countries, the second wave of the pandemic was more severe than the first wave (*5*).

The first cases of COVID-19 on the African continent were reported in Nigeria, Egypt, and South Africa between mid-February and early March 2020, and most countries had reported cases by the end of March 2020 (*6*–*8*). These early cases were concentrated among airline travelers returning from regions of the world with high levels of community transmission. Many African countries introduced early public health and social measures, including international travel controls, quarantine for returning travelers, and internal lockdown measures, to limit the spread of the virus and give health services time to prepare (*5*, 9). The initial phase of the epidemic was then heterogeneous, with relatively high case numbers reported in North Africa and southern Africa, and fewer cases reported in other regions.

From the onset of the pandemic, genomic surveillance has been at the forefront of the COVID-19 response in Africa (*10*). Rapid implementation of SARS-CoV-2 sequencing by various laboratories in Africa enabled genomic data to be generated and shared from the early imported cases. In Nigeria, the first genome sequence was released just 3 days after the announcement of the first case (*6*). Similarly, in Uganda, a sequencing program was set up rapidly to facilitate virus tracing, and the collection of samples for sequencing began immediately upon confirmation of the first case (*11*). In South Africa, the Network for Genomic Surveillance in South Africa (NGS-SA) was established in March 2020, and within weeks, genomic analysis was helping to characterize outbreaks and community transmission (*12*).

Genomic surveillance has also been critical for monitoring ongoing SARS-CoV-2 evolution and detection of new SARS-CoV-2 variants in Africa. Intensified sampling by NGS-SA in the Eastern Cape Province of South Africa in November 2020, in response to a rapid resurgence of cases, led to the detection of B.1.351 (501Y.V2) (*13*). This variant was subsequently designated a variant of concern (VOC) by the World Health Organization (WHO), owing to evidence of increased transmissibility (*14*) and resistance to neutralizing antibodies elicited by natural infection and vaccines (*15*–*17*).

In this study, we performed phylogenetic and phylogeographic analyses of SARS-CoV-2 genomic data from 33 African countries and two overseas territories to help characterize the dynamics of the pandemic in Africa. We show that the early introductions were predominantly from Europe, but that as the pandemic progressed, there was increasing spread between African countries. We also describe the emergence and spread of a number of key SARS-CoV-2 variants in Africa and highlight how the spread of B.1.351 (501Y.V2) and other variants contributed to the more severe second wave of the pandemic in many countries.

All author affiliations are listed at the end of this paper.
*Corresponding author. Email: tulio@sun.ac.za
†These authors contributed equally to this work.

SARS-CoV-2 genomic data

By 5 May 2021, 14,504 SARS-CoV-2 genomes had been submitted to the GISAID database (*18*) from 38 African countries and two overseas territories (Mayotte and Réunion) (Fig. 1A). Overall, this corresponds to approximately one sequence per ~300 reported cases. Almost half of the sequences were from South Africa (n = 5362), consistent with it being responsible for almost half of the reported cases in Africa. Overall, the number of sequences correlates closely with the number of reported cases per country (Fig. 1B). The countries and territories with the highest coverage of sequencing (defined as genomes per reported case) are Kenya (n = 856, one sequence per ~203 cases), Mayotte (n = 721, one sequence per ~21 cases), and Nigeria (n = 660, one sequence per ~250 cases). Although genomic surveillance started early in many countries, few have evidence of consistent sampling across the whole year. Half of all African genomes were deposited in the first 10 weeks of 2021, suggesting intensified surveillance in the second wave after the detection of B.1.351 (501Y.V2) and other variants (Fig. 1, C and D).

Genetic diversity and lineage dynamics in Africa

Of the 10,326 genomes retrieved from GISAID by the end of March 2021, 8746 genomes passed quality control and met the minimum metadata requirements. These genomes from Africa were compared in a phylogenetic framework with 11,891 representative genomes from around the world. Ancestral location state reconstruction of the dated phylogeny (hereafter referred to as discrete phylogeographic reconstruction) allowed us to infer the number of viral imports and exports between Africa and the rest of the world, and between individual African countries. African genomes in this study spanned the whole global genetic diversity of SARS-CoV-2, a pattern that largely reflects multiple introductions over time from the rest of the world (Fig. 2A).

In total, we detected at least 757 [95% confidence interval (CI): 728 to 786] viral introductions into African countries between the start of 2020 and February 2021, more than half of which occurred before the end of May 2020. Although the early phase of the pandemic was dominated by importations from outside Africa, predominantly from Europe, there was then a shift in the dynamics, with an increasing number of importations from other African countries as the pandemic progressed (Fig. 2, B and C). A rarefaction analysis in which we systematically subsampled genomes shows that vastly more introductions would have likely been identified with increased sampling in Africa or globally, suggesting that the introductions we identified are really just the "ears of the hippo," or a small part of a larger problem (fig. S1).

South Africa, Kenya, and Nigeria appear as major sources of importations into other African countries (Fig. 2D), although this is likely to be influenced by these three countries having the greatest number of deposited sequences. Particularly notable is the southern African region, where South Africa is the source for a large proportion (~80%) of the importations to other countries in the region. The North African region demonstrates a different pattern to the rest of the continent, with more viral introductions from Europe and Asia (particularly the Middle East) than from other African countries (fig. S2).

Africa has also contributed to the international spread of the virus, with at least 324 (95% CI: 295 to 353) exportation events from Africa to the rest of the world detected in this dataset. Consistent with the source of importations, most exports were to Europe (41%), Asia (26%), and North America (14%). As with the number of importations, exports were relatively evenly distributed over the 1-year period (fig. S3). However, an increase in the number of exportation events occurred between December 2020 and March 2021, which coincided with the second wave of infections in Africa and with some relaxations of travel restrictions around the world.

The early phase of the pandemic was characterized by the predominance of lineage B.1. This was introduced multiple times to African countries and has been detected in all but one of the countries included in this analysis. After its emergence in South Africa, B.1.351 became the most frequently detected SARS-CoV-2 lineage found in Africa (n = 1769, ~20%) (Fig. 1C). It was first sampled on 8 October 2020 in South Africa (*13*) and has since spread to 20 other African countries.

As air travel came to an almost complete halt in March and April 2020, the number(s) of detectable viral imports into Africa decreased and the pandemic entered a phase that was characterized in sub-Saharan Africa by sustained low levels of within-country movements and occasional international viral movements between neighboring countries, presumably via road and rail links between these. Though some border posts between countries were closed during the initial lockdown period (table S1), others remained open to allow trade to continue. Regional trade in southern Africa was only slightly affected by lockdown restrictions and quickly rebounded to prepandemic levels (fig. S4) after the relaxation of restrictions between June 2020 and December 2020.

Although lineage A viruses were imported into several African countries, they only account for 1.3% of genomes sampled in Africa. Despite lineage A viruses initially causing many localized clustered outbreaks, each the result of independent introductions to several countries (e.g., Burkina Faso, Côte d'Ivoire, and Nigeria), they were later largely replaced by lineage B viruses as the pandemic evolved. This is possibly due to the increased transmissibility of lineage B viruses by virtue of the D614G (Asp^{614}→Gly) mutation in the spike protein (*19*, *20*). However, there is evidence of an increasing prevalence of lineage A viruses in some African countries (*11*). In particular, A.23.1 emerged in East Africa and appears to be rapidly increasing in prevalence in Uganda and Rwanda (*11*). Furthermore, a highly divergent variant from lineage A was recently identified in Angola from individuals arriving from Tanzania (*21*).

Emergence and spread of new SARS-CoV-2 variants

To determine how some of the key SARS-CoV-2 variants are spreading within Africa, we performed phylogeographic analyses on the VOC B.1.351, the variant of interest (VOI) B.1.525, and two additional variants that emerged and that we designated as VOIs for this analysis (A.23.1 and C.1.1). These African VOCs and VOIs have multiple mutations on the spike glycoprotein, and a molecular clock analysis of these four datasets provided strong evidence that these four lineages are evolving in a clock-like manner (Fig. 3, A and B).

B.1.351 was first sampled in South Africa in October 2020, but phylogeographic analysis suggests that it emerged earlier, around August 2020. It is defined by 10 mutations in the spike protein, including K417N (Lys^{417}→Asn), E484K (Glu^{484}→Lys), and N501Y (Asn^{501}→Tyr) in the receptor binding domain (Fig. 3B). After its emergence in the Eastern Cape, it spread extensively within South Africa (Fig. 4A). By November 2020, the variant had spread into neighboring Botswana and Mozambique, and by December 2020, it had reached Zambia and Mayotte. Within the first 3 months of 2021,

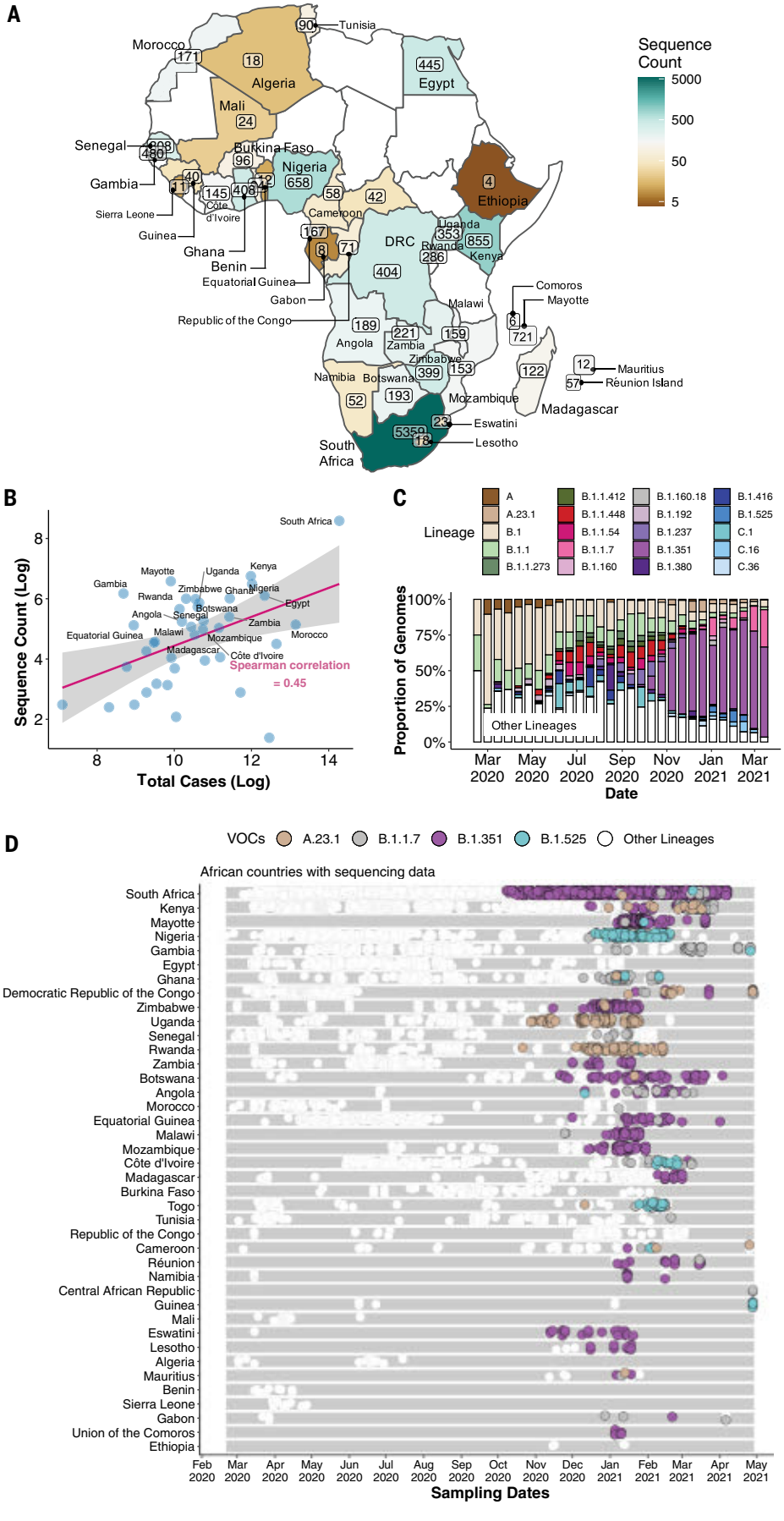


Fig. 1. SARS-CoV-2 sequences in Africa. (**A**) Map of the African continent with the number of SARS-CoV-2 sequences reflected in GISAID as of 5 May 2021. (**B**) Regression plot of the number of viral sequences versus the number of reported COVID-19 cases in various African countries as of 5 May 2021. Countries with >500 sequences are labeled. The shaded region indicates the 95% confidence interval. (**C**) Progressive distribution of the top 20 PANGO lineages on the African continent. (**D**) Temporal sampling of SARS-CoV-2 sequences in African countries (ordered by total number of sequences) through time, with VOCs of note highlighted and annotated according to their PANGO lineage assignment.

further exports from South Africa into Botswana, Zimbabwe, Mozambique, and Zambia occurred. By March 2021, B.1.351 had become the dominant lineage within most southern African countries as well as the overseas territories of Mayotte and Réunion (fig. S5). Our phylogeographic reconstruction also demonstrates movement of B.1.351 into East and Central Africa directly from southern Africa. Our discrete phylogeographic analysis of a wider sample of B.1.351 isolates demonstrates the spread of the lineage into West Africa. This patient from West Africa had a known travel history to Europe, so it is possible that the patient acquired the infection while in Europe or in transit and not from other African sources (fig. S6).

B.1.525 is a VOI defined by six substitutions in the spike protein [Q52R (Gln^{52}→Arg), A67V (Ala^{67}→Val), E484K, D614G, Q677H (Gln^{677}→His), and F888L (Phe^{888}→Leu)] and two deletions in the N-terminal domain [HV69-70Δ (deletion of His and Val at positions 69 and 70) and Y144Δ (deletion of Tyr at position 144)]. This was first sampled in the United Kingdom in mid-December 2020, but our phylogeographic reconstruction suggests that the variant originated in Nigeria in November 2020 [95% highest posterior density (HPD) 2020-11-01 to 2020-12-03] (Fig. 4B). Since then, it has spread throughout much of Nigeria and neighboring Ghana. Given sparse sampling from other neighboring countries within West and Central Africa (Fig. 1, A and C), the extent of the spread of this VOI in the region is not clear. Beyond Africa, this VOI has spread to Europe and the United States (fig. S6).

We designated A.23.1 and C.1.1 as VOIs for the purposes of this analysis because they present good examples of the continued evolution of the virus within Africa (*11*, *13*). Lineage A.23, characterized by three spike mutations [F157L (Phe^{157}→Leu), V367F (Val^{367}→Phe), and Q613H (Gln^{613}→His)], was first detected in a Ugandan prison in Amuru in July 2020 (95% HPD: 2020-07-15 to 2020-08-02). From there, the lineage was transmitted to Kitgum prison, possibly facilitated by the transfer of prisoners. Subsequently, the A.23 lineage spilled into the

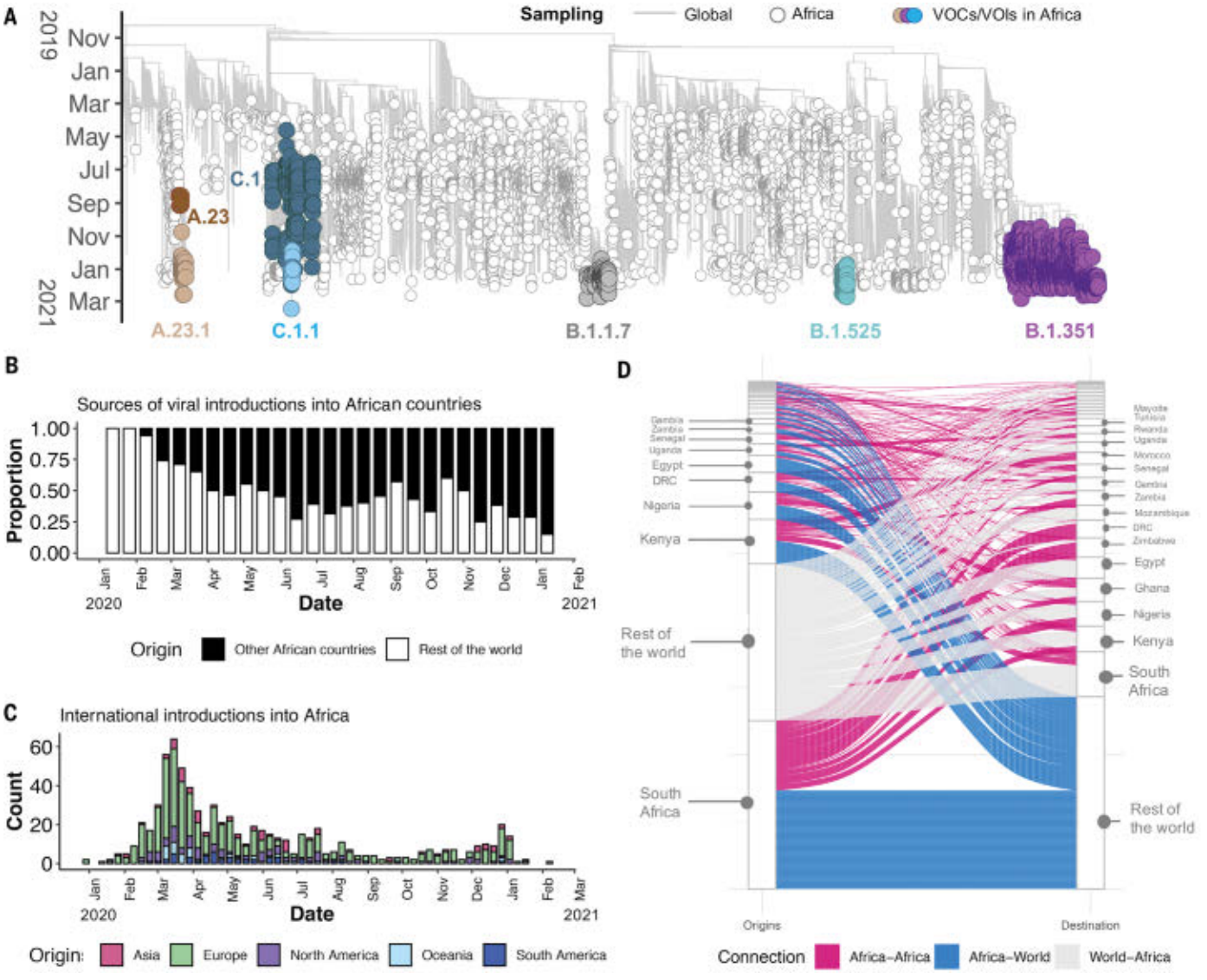


Fig. 2. Phylogenetic reconstruction of the SARS-CoV-2 pandemic on the continent of Africa. (**A**) Time-resolved maximum likelihood tree containing 8746 high-quality African SARS-CoV-2 near-full-genome sequences analyzed against a backdrop of global reference sequences. VOIs and VOCs are highlighted on the phylogeny. (**B**) Sources of viral introductions into African countries characterized as external introductions from the rest of the world versus internal introductions from other African countries. (**C**) Total external viral introductions over time into Africa. (**D**) The number of viral imports and exports into and out of various African countries depicted as internal (between African countries, in pink) or external (between African and non-African countries, in blue and gray).

general population and spread to Kampala, adding other spike mutations [R102I (Arg^{102}→Ile), L141F (Leu^{141}→Phe), E484K, and P681R (Pro^{681}→Arg)] along with additional mutations in nsp3, nsp6, ORF8, and ORF9, prompting a new lineage classification, A.23.1 (Fig. 3, A and B). Since the emergence of A.23.1 in September 2020 (95% HPD: 2020-09-02 to 2020-09-28), it has spread regionally into neighboring Rwanda and Kenya and has now also reached South Africa and Botswana in the south and Ghana in the west (Fig. 4C). However, our phylogeographic reconstruction of A.23.1 suggests that the introduction into Ghana may have occurred via Europe (fig. S6), whereas the introductions

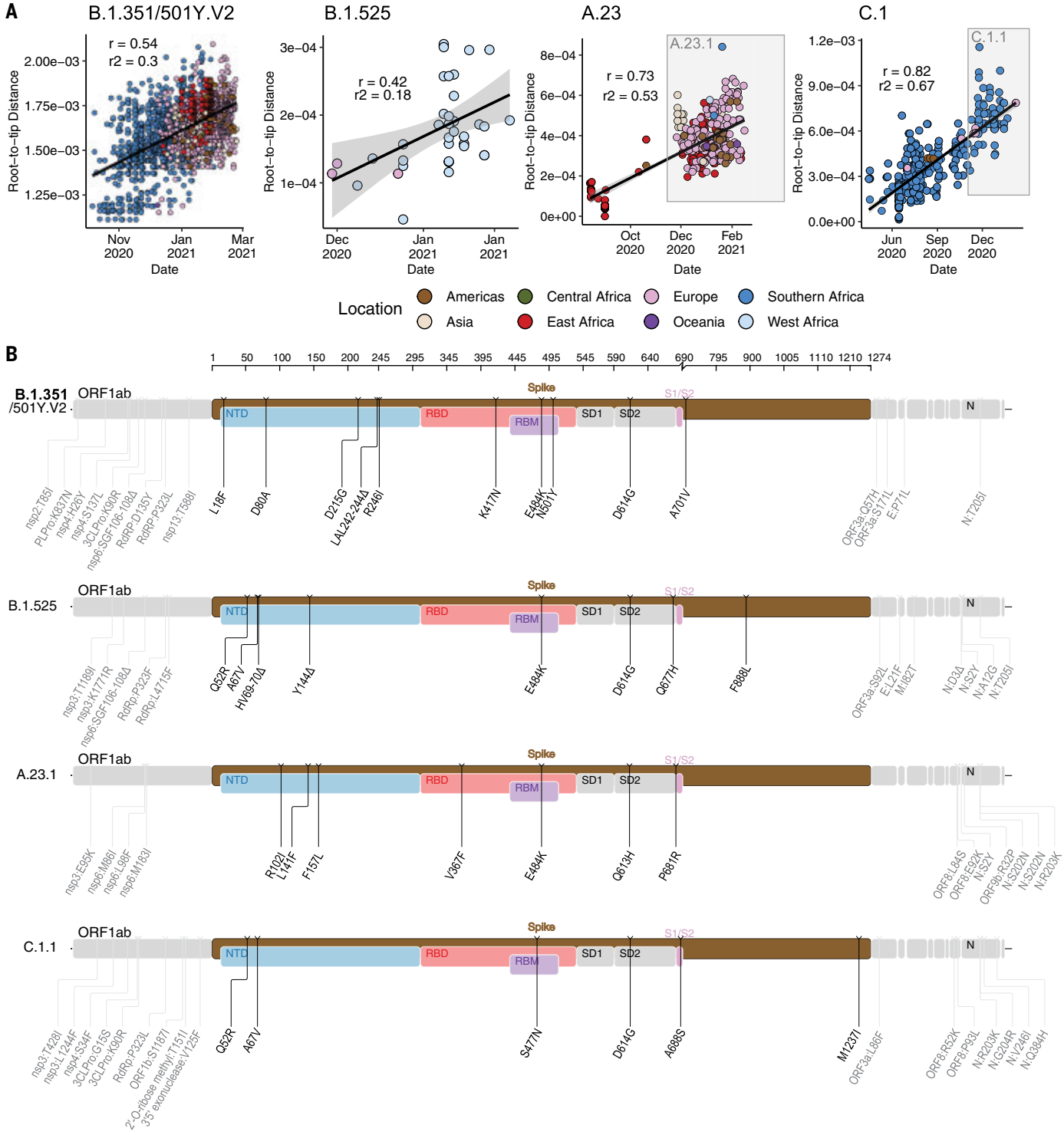


Fig. 3. Genetic profile of VOCs and VOIs under investigation. (**A**) Root-to-tip regression plots for four lineages of interest. C.1 and A.23 show continued evolution into VOIs C.1.1 and A.23.1, respectively. *r*, coefficient of correlation; r^2, coefficient of determination. (**B**) Genome maps of four VOCs and VOIs, where the spike region is shown in detail and in color and the rest of the genome is shown in gray. ORF, open reading frame; NTD, N-terminal domain; RBD, receptor binding domain; RBM, receptor binding motif; SD1, subdomain 1; SD2, subdomain 2.

into southern Africa likely occurred directly from East Africa. This is consistent with epidemiological data suggesting that the case detected in South Africa was a contact of an individual who had recently traveled to Kenya.

Lineage C.1 emerged in South Africa in March 2020 (95% HPD: 2020-03-13 to 2020-04-17) during a cluster outbreak before the first wave of the epidemic (*13*). C.1.1 is defined by the spike mutations S477N (Ser^{477}→Asn), A688S (Ala^{688}→Ser), and M1237I (Met^{1237}→Ile) and also contains the Q52R and A67V mutations similar to B.1.525 (Fig. 3B). A continuous trait phylogeographic reconstruction of the movement dynamics of these lineages suggests that C.1 emerged in the city of Johannesburg and spread within South Africa during the first wave (Fig. 4D). Independent exports of C.1 from South Africa led to regional spread to Zambia (June to July 2020) and Mozambique (July to August 2020), and the evolution to C.1.1 seems to have occurred in Mozambique around mid-September 2020 (95% HPD: 2020-09-07 to 2020-10-05). An in-depth analysis of SARS-CoV-2 genotypes from Mozambique suggests that the C.1.1 lineage was the most prevalent in the country until the introduction of B.1.351, which has dominated the epidemic since (fig. S5).

The VOC B.1.1.7, which was first sampled in Kent, England, in September 2020 (*22*), has also increased in prevalence in several African countries (fig. S5). To date, this VOC has been detected in 11 African countries, as well as the Indian Ocean islands of Mauritius and Mayotte (fig. S7). The time-resolved phylogeny suggests that this lineage was introduced into Africa on at least 16 occasions between November 2020 and February 2021, with evidence of local transmission in Nigeria and Ghana.

Conclusions

Our phylogeographic reconstruction of past viral dissemination patterns suggests a strong epidemiological linkage between Europe and Africa, with 64% of detectable viral imports into Africa originating in Europe and 41% of detectable viral exports from Africa landing in Europe (Fig. 1C). This phylogeographic analysis also

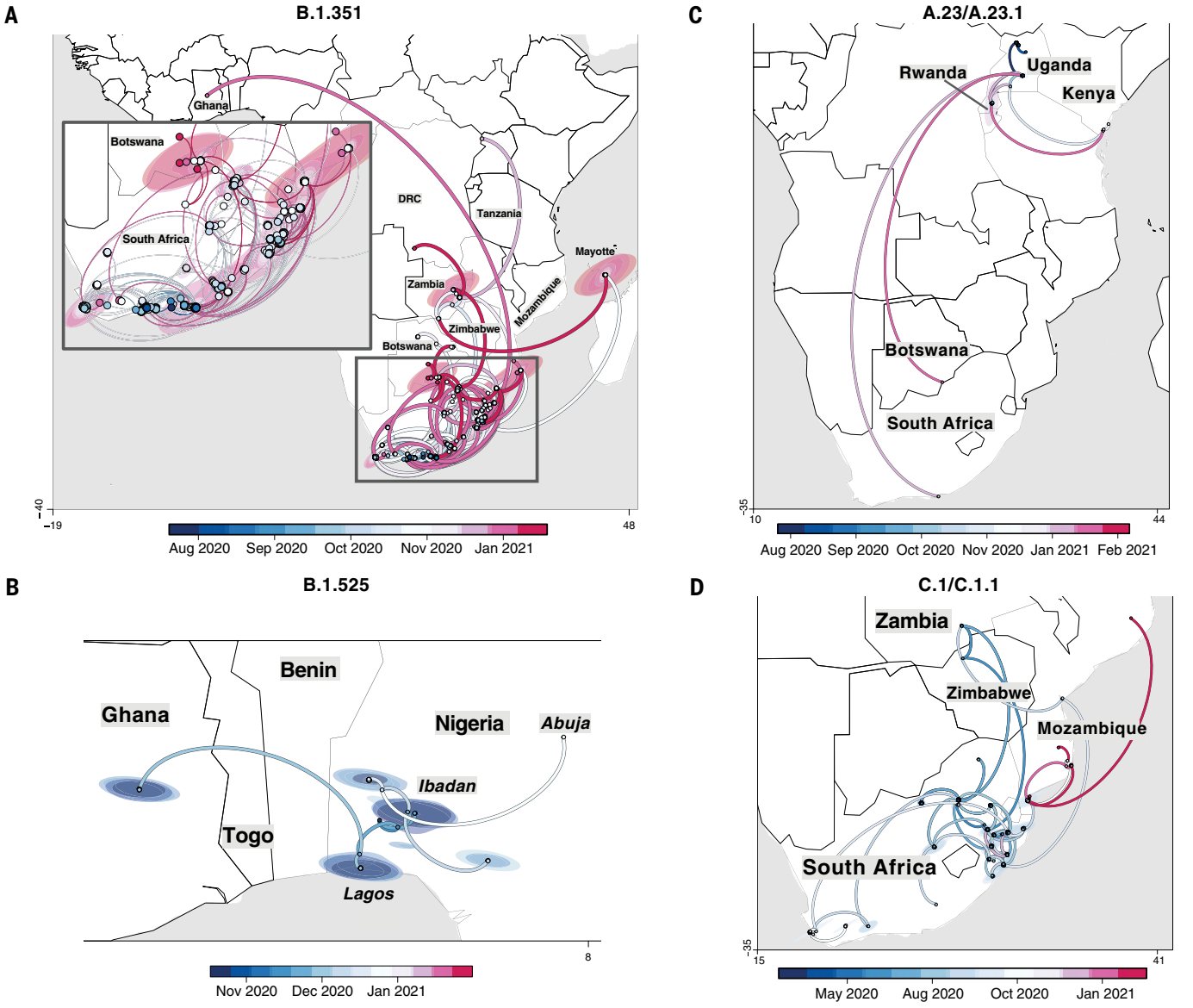


Fig. 4. Phylogeographic reconstruction of the spread of four VOCs and VOIs across the African continent. (**A** to **D**) Phylogeographic reconstruction of the spread of four VOCs and VOIs across the African continent using sequences showing strict continuous transmission across geographical regions: B.1.351 (A), B.1.525 (B), A.23/A.23.1 (C), and C.1/C.1.1 (D). Curved lines denote the direction of transmission in the counterclockwise direction. Solid lines show transmission paths as inferred by phylogeographic reconstruction and colored by date, whereas dashed lines show the known travel history of the particular case considered.

suggests a changing pattern of viral diffusion into and within Africa over the course of 2020. In almost all instances, the earliest introductions of SARS-CoV-2 into individual African countries were from countries outside Africa.

High rates of COVID-19 testing and consistent genomic surveillance in the south of the continent have led to the early identification of VOCs such as B.1.351 and VOIs such as C.1.1 (*13*). Since the discovery of these southern African variants, several other SARS-CoV-2 VOIs have emerged in different parts of the world, including elsewhere on the African continent, such as B.1.525 in West Africa and A.23.1 in East Africa. There is strong evidence that both of these VOIs are rising in frequency in the regions where they have been detected, which suggests that they may possess higher fitness than other variants in these regions. Although more-focused research on the biological properties of these VOIs is needed to confirm whether they should be considered VOCs, it would be prudent to assume the worst and focus on limiting their spread. It will be important to investigate how these different variants compete against one another if they occupy the same region.

Our focused phylogenetic analysis of the B.1.351 lineage revealed that in the final months of 2020, this variant spread from South Africa into neighboring countries, reaching as far north as the Democratic Republic of the Congo (DRC) by February 2021. This spread may have been facilitated through rail and road networks that form major transport arteries linking South Africa's ocean ports to commercial and industrial centres in Botswana, Zimbabwe, Zambia, and the southern parts of the DRC. The rapid, apparently unimpeded spread of B.1.351 into these countries suggests that current land-border controls that are intended to curb the international spread of the virus are ineffective. Perhaps targeted testing of cross-border travelers, genotyping of positive cases, and the focused tracking of frequent cross-border travelers, such as long distance truckers, would more effectively contain the spread of future VOCs and VOIs that emerge within this region.

The dominance of VOIs and VOCs in Africa has important implications for vaccine rollouts on the continent. For one, slow rollout of vaccines in most African countries creates an environment in which the virus can replicate and evolve: This will almost certainly produce additional VOCs, any of which could derail the global fight against COVID-19. Conversely, with the already widespread presence of known variants, difficult decisions about balancing reduced efficacy and availability of vaccines have to be made. This also highlights how crucial it is that trials are done. From a public health perspective, genomic surveillance is only one item in the toolkit of pandemic preparedness. It is important that such work is closely followed by genotype-to-phenotype research to determine the actual relevance of continued evolution of SARS-CoV-2 and other emerging pathogens.

The rollout of vaccines across Africa has been painfully slow (figs. S8 and S9). There have, however, been notable successes that suggest that the situation is not hopeless. The small island nation of the Seychelles had vaccinated 70% of its population by May 2021. Morocco has kept pace with many developed nations and, by mid-March, had vaccinated ~16% of its population. Rwanda, one of Africa's most resource-constrained countries, had, within 3 weeks of obtaining its first vaccine doses in early March, managed to provide first doses to ~2.5% of its population. For all other African countries, at the time of writing, vaccine coverage (first dose) was <1.0% of the general population.

The effectiveness of molecular surveillance as a tool for monitoring pandemics is largely dependent on continuous and consistent sampling through time, rapid virus genome sequencing, and rapid reporting. When this is achieved, molecular surveillance can ensure the early detection of changing pandemic characteristics. Further, when such changes are discovered, molecular surveillance data can also guide public health responses. In this regard, the molecular surveillance data that are being gathered by most African countries are less useful than they could be. For example, the time lag between when virus samples are taken and when sequences for these samples are deposited in sequence repositories is so great in some cases that the primary utility of genomic surveillance data is lost (fig. S10). This lag is driven by several factors, depending on the laboratory or country in question: (i) lack of reagents owing to disruptions in global supply chains, (ii) lack of equipment and infrastructure within the originating country, (iii) scarcity of technical skills in laboratory methods or bioinformatic support, and (iv) hesitancy by some health officials to release data. More-recent sampling and prompt reporting is crucial to reveal the genetic characteristics of currently circulating viruses in these countries.

The patchiness of African genomic surveillance data is therefore the main weakness of our study. However, there is evidence that the situation is improving, with ~50% of African SARS-CoV-2 genome sequences having been submitted to the GISAID database within the first 10 weeks of 2021. Although the precise factors underlying this surge in sequencing efforts are unclear, an important driver is almost certainly increased global interest in genomic surveillance after the discovery of multiple VOCs and VOIs since December 2020. We cannot reject that the observed increase in exports from Africa may be due to intensified sequencing activity after the detection of variants around the world. It is important to note here that phylogeographic reconstruction of viral spread is highly dependent on sampling where there is the caveat that the exact routes of viral movements between countries cannot be inferred if there is no sampling in connecting countries. Furthermore, our efforts to reconstruct the movement dynamics of SARS-CoV-2 across the continent are almost certainly biased by uneven sampling between different African countries. It is not a coincidence that we identified South Africa, Kenya, and Nigeria, which have sampled and sequenced the most SARS-CoV-2 genomes, as major sources of viral transmissions between sub-Saharan African countries. However, these countries also had the highest number of infections, which may decrease the sampling biases (Fig. 1A).

The reliability of genomic surveillance as a tool to prevent the emergence and spread of dangerous variants is dependent on the intensity with which it is embraced by national public health programs. As with most other parts of the world, the success of genomic surveillance in Africa requires that more samples are tested for COVID-19, higher proportions of positive samples are sequenced within days of sampling, and persistent analyses of these sequences are performed for concerning signals such as (i) the presence of novel nonsynonymous mutations at genomic sites associated with pathogenicity and immunogenicity, (ii) evidence of positive selection at codon sites where nonsynonymous mutations are observed, and (iii) evidence of lineage expansions. Despite limited sampling, Africa has identified many of the VOCs and VOIs that are being transmitted across the world. Detailed characterization of the variants and their impact on vaccine-induced immunity is of extreme importance. If the pandemic is not controlled in Africa, we may see the production of vaccine escape variants that may profoundly affect the population in Africa and across the world.

REFERENCES AND NOTES

1. C. Wang, P. W. Horby, F. G. Hayden, G. F. Gao, *Lancet* **395**, 470–473 (2020).
2. Q. Li *et al*., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
3. S. Uyoga *et al*., *Science* **371**, 79–82 (2021).
4. L. Mwananyanda *et al*., *BMJ* **372**, n334 (2021).
5. S. J. Salyer *et al*., *Lancet* **397**, 1265–1275 (2021).
6. P. Oluniyi, First African SARS-CoV-2 genome sequence from Nigerian COVID-19 case. Virological (2020); https://virological.org/t/first-african-sars-cov-2-genome-sequence-from-nigerian-covid-19-case/421.
7. M. A. Medhat, M. El Kassas, *J. Glob. Health* **10**, 010368 (2020).
8. M. Allam *et al*., *Microbiol. Resour. Announc.* **9**, e00572-e20 (2020).
9. N. Haider *et al*., *BMJ Glob. Health* **5**, e003319 (2020).
10. S. C. Inzaule, S. K. Tessema, Y. Kebede, A. E. Ogwell Ouma, J. N. Nkengasong, *Lancet Infect. Dis.* **21**, e281–e289 (2021).
11. D. L. Bugembe *et al*., *medRxiv* 2021.02.08.21251393 (2021); https://doi.org/10.1101/2021.02.08.21251393.
12. J. Giandhari *et al*., *Int. J. Infect. Dis.* **103**, 234–241 (2021).
13. H. Tegally *et al*., *Nat. Med.* **27**, 440–446 (2021).
14. C. A. Pearson, T. W. Russell, N. Davies, A. J. Kucharski, Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa. CMMID Repository (2021); https://cmmid.github.io/topics/covid19/sa-novel-variant.html.

15. S. Cele *et al*., *medRxiv* 2021.01.26.21250224 (2021); https://doi.org/10.1101/2021.01.26.21250224.
16. S. A. Madhi *et al*., *N. Engl. J. Med.* **384**, 1885–1898 (2021).
17. C. K. Wibmer *et al*., *Nat. Med.* **27**, 622–625 (2021).
18. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
19. E. Volz *et al*., *Cell* **184**, 64–75.e11 (2021).
20. B. Korber *et al*., *Cell* **182**, 812–827.e19 (2020).
21. T. de Oliveira *et al*., *medRxiv* 2021.03.30.21254323 (2021); https://doi.org/10.1101/2021.03.30.21254323.
22. B. Meng *et al*., *Cell Rep.* **35**, 109292 (2021).
23. S. E. James, T. Houriiyah, krisp-kwazulu-natal/africa-covid19-genomics: A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa - Code and scripts. Zenodo (2021); https://doi.org/10.5281/zenodo.5386379.

ACKNOWLEDGMENTS

We acknowledge the authors from the originating laboratories and the submitting laboratories, who generated and shared, via GISAID, the genetic sequence data on which this research is based (table S4). We also acknowledge the contribution of K. Maria from the NGS-SA platform for their contribution toward the sequencing effort in Cape Town, South Africa. Similarly, we thank A. M. Elsaame, S. M. Elsayed, and R. M. Darwish from the Faculty of Medicine Ain Shams Research Institute (MASRI) for their efforts toward sequencing in Egypt. We thank S. Bane, M. Sanogo, D. Diallo, A. Combo Georges Togo, and A. Coulibaly from the University Clinical Research Centre (UCRC) at the University of Sciences, Techniques, and Technologies of Bamako for the contribution they have made toward sequencing efforts in Mali. We acknowledge the contribution of M. Moeti and A. Salam Gueye from the WHO for their contribution toward combating SARS-CoV-2 on the African continent. We further wish to extend acknowledgment to S. Lutucuta and J. Morais from the Angolan Ministry of Health for their continued hard work with regards to SARS-CoV-2 sampling, sequencing, and pandemic response in Angola. From Malawi we wish to acknowledge the work of B. Chilima, B. Mvula, and M. Chitenje from the Malawian Ministry of Health for their work on the COVID-19 response within the country. **Funding:** The University of Ghana (WACCBIP) team was funded by a Wellcome/African Academy of Sciences Developing Excellence in Leadership Training and Science (DELTAS) grant (DEL-15-007 and 107755/Z/15/Z: Awandare); National Institute of Health Research (NIHR) (17.63.91) grants using UK aid from the UK government for a global health research group for genomic surveillance of malaria in West Africa (Wellcome Sanger Institute, UK) and the global research unit for Tackling Infections to Benefit Africa (TIBA partnership, University of Edinburgh); and a World Bank African Centres of Excellent grant (WACCBIP-NCDs: Awandare). Project ADAGE PRFCOV19-GP2 (2020-2022) includes 40 researchers from the Center of Biotechnology of Sfax, the University of Sfax, the University of Monastir, the University Hospital Hédi Chaker of Sfax, the Military Hospital of Tunis, and Dacima Consulting. Ministry of Higher Education and Scientific Research and Ministry of Health of the Republic of Tunisia. The Uganda contributions were funded by the UK Medical Research Council (MRC/UKRI) and the UK Department for International Development (DFID) under the MRC/DFID concordat agreement (grant agreement number NC_PC_19060) and by the Wellcome, DFID–Wellcome Epidemic Preparedness–Coronavirus (grant agreement number 220977/Z/20/Z) awarded to M.C. Work from Quadram Institute Bioscience was funded by The Biotechnology and Biological Sciences Research Council Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10351, and BBS/E/F/000PR10352 and by the Quadram Institute Bioscience BBSRC–funded Core Capability Grant (project number BB/CCG1860/1). The Africa Pathogen Genomics Initiative (Africa PGI) at the Africa CDC is supported by the Bill & Melinda Gates Foundation (INV018978 and INV018278), Illumina Inc, the US Centers for Disease Control and Prevention (CDC), and Oxford Nanopore Technologies. Sequences generated in Zambia through PATH were funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation. Funding for sequencing in Côte d'Ivoire, Burkina Faso, and part of the sequencing in the DRC was granted by the German Federal Ministry of Education and Research (BMBF). Sequencing efforts from Morocco have been supported by Academie Hassan II of Science and Technology, Morocco. Funding for surveillance, sampling, and testing in Madagasar was provided by the WHO, the CDC (grant U5/IP000812-05), the US Agency for International Development (USAID; cooperation agreement 72068719CA00001), and the Office of the Assistant Secretary for Preparedness and Response in the US Department of Health and Human Services (DHHS; grant number IDSEP190051-01-0200). Funding for sequencing was provided by the Bill & Melinda Gates Foundation (GCE/ID OPP1211841), Chan Zuckerberg Biohub, and the Innovative Genomics Institute at UC Berkeley. The Botswana Harvard AIDS Institute was supported by the following funding: H3ABioNet through funding from the National Institutes of Health Common Fund (U41HG006941)—H3ABioNet is an initiative of the Human Health and Heredity in Africa Consortium (H3Africa) program of the African Academy of Science (AAS); DHHS–NIH–National Institute of Allergy and Infectious Diseases (NIAID) (5K24AI131928-04 and 5K24AI131924-04); Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE); a DELTAS Africa Initiative (grant DEL-15-0060—the DELTAS Africa Initiative is an independent funding scheme of the AAS's Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107752/Z/15/Z] and the UK government; and the South African Medical Research Council (SAMRC) and the Department of Technology and Innovation as part of the Network for Genomic Surveillance in South Africa (NGS-SA) and the Stellenbosch University Faculty of Medicine & Health Sciences, Strategic Equipment Fund. D.P.M. is funded by the Wellcome Trust (Wellcome Trust grant 222574/Z/21/Z). Sequencing activities at the NICD were supported by a conditional grant from the South African National Department of Health as part of the emergency COVID-19 response; a cooperative agreement between the National Institute for Communicable Diseases of the National Health Laboratory Service and the U.S. Centers for Disease Control and Prevention (grant number 5 U01IP001048-05-00); the African Society of Laboratory Medicine (ASLM) and Africa Centers for Disease Control and Prevention through a sub-award from the Bill and Melinda Gates Foundation grant number INV-018978; the UK Foreign, Commonwealth and Development Office and Wellcome (Grant no 221003/Z/20/Z); the South African Medical Research Council (Reference number SHIPNCD 76756); the UK Department of Health and Social Care, managed by the Fleming Fund and performed under the auspices of the SEQAFRICA project. Furthermore, pandemic surveillance in South Africa and Senegal was supported in part through NIH grant U01 AI151698 for the United World Antiviral Research Network (UWARN). Support for pandemic surveillance from the Tulio de Oliveira group to other African countries is funded by the Rockefeller Foundation. Sequencing efforts in the DRC were funded by the Bill & Melinda Gates Foundation under grant INV-018030 awarded to C.B.P. and further supported by funding from the Africa CDC through the ASLM (African Society of Laboratory Medicine) for Accelerating SARS-CoV-2 Genomic Surveillance in Africa. Sequencing efforts in Rwanda were commissioned by the NIHR Global Health Research program (16/136/33) using UK aid from the UK government (funding to E.M. and N. R. through TIBA partnership) and additional funds from the government of Rwanda through RBC/National Reference Laboratory in collaboration with the Belgian Development Agency (ENABEL) for additional genomic sequencing at GIGA Research Institute–Liege/Belgium. The sequencing effort in Equatorial Guinea was supported by a public-private partnership, the Bioko Island Malaria Elimination Project, composed of the government of Equatorial Guinea Ministries of Mines and Hydrocarbons, and Health and Social Welfare, Marathon EG Production Limited, Noble Energy, Atlantic Methanol Production Company, and EG LNG. Sample collection and typing in Mali were supported by Fondation Merieux–France, and sequence efforts have been supported by the Enable and Enhance Initiative of the German Federal Government's Security Cooperation against Biological Threats in the G5 Sahel Region. The Nigeria work was made possible by support from Flu Lab and a cohort of generous donors through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott, the Skoll Foundation, and Open Philanthropy. Further Nigeria funding came from grants from the NIAID (www.niaid.nih.gov), NIH-H3Africa (https://h3africa.org) (U01HG007480 and U54HG007480), and the World Bank grant (worldbank.org) (ACE IMPACT project) to C.H. Analysis for the Gabon strains was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan International Cooperation Agency (JICA), and Japan Agency for Medical Research and Development (AMED) (grant number JP20jm0110013) and a grant from AMED (grant number JP20wm0225003). Sequencing at KEMRI-Wellcome Trust Research Programme site in Kenya was supported by the National Institute for Health Research (NIHR) (project references 17/63/82 and 16/136/33), using UK aid from the UK Government to support global health research, and the UK Foreign, Commonwealth and Development Office and Wellcome Trust (grant# 102975; 220985). **Author contributions:** Conceptualization: E.W., H.T., J.G.S., J.O., J.O.G., K.O.D., R.A.D., R.A.K., R.L., S.K.T., S.Ma., T.d.O. Methodology: A.-R.N.Z., A.Re., C.N.A., D.P.M., D.A.R., E.W., H.T., J.A.E., J.Gy., J.G.S., K.H.Y., K.O.D., L.d.O.M., M.A.B., M.C., M.G., M.M.N., M.V.T.P., P.A., T.d.O., V.F. Investigation: A.E., A.Ir., A.-R.N.Z., A.Re., A.So., A.v.G., C.A.K., C.W., D.C., D.J.N., D.P.M., D.A.R., D.S., E.M., E.N.N., E.W., F.L., G.G., H.T., J.A.E., J.E.S., J.Gy., J.G.S., J.-J.M.T., J.L., J.Nk., J.Q.M., K.H.Y., M.-M.D., M.A.B., M.C., M.G., M.Mar., M.Mas., M.S., M.V.T.P., N.A., N.H.R., N.K., P.K., R.A.A.C., R.A.D., R.G., S.A.M., S.F.A., S.Ma., S.O., T.L.V., V.F., W.P. Sampling: A.-S.K., A.N.A., A.A.A., A.A.S., A.D., A.E., A.E.O.O., A.F., A.G., A.H., A.Ib., A.Ka., A.K.Sa., A.K.Se., A.L., A.M.O., A.O.O.O., A.P., A.Ro., A.Sy., A.F.V., A.V.G., A.v.G., B.B., B.L.H., B.Ko., B.Kl., B.Ma., B.N., B.S.O., B.T., C.N.A., C.D., C.P., C.S., C.W., D.Bas., D.C., D.D., D.G., D.G.A., D.J.N., D.S., E.K.L., E.M., E.M.O., E.N.N., E.P., E.Sh., F.Ab., F.Ad., F.O.E., F.M.M., F.S.M., F.O., F.T., F.T.T., G.A.A., G.G., G.K.M., G.P.M., G.T., G.v.Z., H.C., H.A.E., H.N., I.C., I.Gaa., I.Gaz., I.K., I.M., I.O., I.Ss., J.C.A., J.-C.M., J.B.L.-D., J.E.S., J.Gi., J.Gy., J.J.L., J.K., J.L., J.M.H., J.M.M., J.Nam., J.Nak., J.T.K., K.T.A., K.M.S., L.B., M.-M.D., M.Al., M.C., M.D., M.e.H., M.G.S., M.K.K., M.Mp., M.Mar., M.Mas., M.M.D., M.N., M.Ou., M.O.A., M.R., M.S., M.W.M., M.T.Y., N.A., N.G., N.H., N.H.R., N.Ig., N.K., N.Ma., N.Nd., N.B.S., N.S., O.Ch., O.E.C., O.Fa., O.Fe., O.I., O.J., O.C.K., O.E.O., O.P., O.S., O.Te., P.E.O., P.A.B., P.C., P.C.S., P.D., P.K., P.K.Q., P.O., P.S., R.A.D., R.G., R.N., S.A., S.S.A., S.B., S.B.L., S.D., S.EI., S.E.K., S.F.A., S.Gas., S.Kam., S.L., S.L.D., S.Me., S.Mo., S.M.M., S.N., S.Pi., S.S., S.To., T.L.V., T.Ma., T.S., U.C., U.G., U.J., U.R., V.G., W.K.A., W.T.C., W.P., W.H.R., Y.Bu., Y.K.T., Y.N., Z.R.D. Sequencing: A.-S.Kam., A.N.A., A.A.A., A.A.S., A.C., A.D., A.E.O.O., A.F., A.Ib., A.Is., A.Ko., A.Ka., A.K.K., A.K.Sa., A.K.Se., A.L., A.-R.N.Z., A.P., A.Sa., A.So., A.Sy., A.S.O., A.J.T., A.F.V., A.V.G., A.v.G., A.A.Y., B.B., B.D., B.L.H., B.Ko., B.Kl., B.Mc., B.N., B.T., C.N.A., C.B., C.B.P., C.D., C.M.M., C.P., C.S., D.Bas., D.D., D.G., D.G.A., D.J.B., D.L.B., D.M., D.O.O., D.P., D.S.Y.A., D.T., E.E.F., E.F.N., E.K.L., E.L., E.M.O., E.P., E.Sh., E.Si., E.S.L., F.Ab., F.Aj., F.A.D., F.D., F.M.M., F.S.M., F.O., F.T.T., F.W., G.A.A., G.G., G.P.M., G.T., G.v.Z., H.Ab., H.An., H.C., H.C.R., H.A.E., H.G., H.H.K., H.N., I.B.-B.B., I.C., I.Gaa., I.Gaz., I.K., I.M., I.Ss., J.C.A., J.B., J.-C.M., J.B.L.-D., J.F., J.Gi., J.Gy., J.J.L., J.K., J.M.H., J.M.M., J.M.N., J.Nam., J.Nak., J.Q.M., J.T.K., J.Y., K.T.A., K.M.S., K.O.D., K.S., K.A.T., L.B., L.F., L.S., L.T., M.-M.D., M.Al., M.A.B., M.C., M.D., M.e.H., M.G.S., M.I.M., M.K.K., M.Mi., M.Mp., M.Mw., M.M.D., M.M.N., M.Ow., M.Ou., M.O.A., M.V.T.P., M.W.M., M.T.Y., N.D., N.G., N.H., N.Ig., N.Is., N.Ma., N.Nd., N.Ns., N.B.S., N.S., N.T., O.Cy., O.Ch., O.E.C., O.Fa., O.I., O.J., O.Te., P.A., P.A.B., P.C.S., P.D., P.E.O., P.K., P.K.Q., P.K.M., P.O., P.S., R.A.A.C., R.G., R.N., R.O.P., S.A., S.S.A., S.B., S.B.L., S.C.S., S.D., S.EI., S.En., S.E.K., S.Gar., S.Gas., S.H.Ab., S.Kas., S.L., S.L.D., S.Me., S.Mo., S.Ma., S.M.M., S.N., S.Pi., S.Pr., S.R., S.S., S.To., S.Tr., S.v.d.W., T.A., T.Ma., T.Mo., T.S., U.C., U.G., U.J., U.J.A., U.R., V.G., W.K.A., W.T.C., W.H.R., Y.Bu., Y.Be., Y.K.T., Y.N., Z.R.D. Visualization: A.C., A.Is., A.Ko., A.K.K., A.Sa., A.So., A.A.Y., B.T., C.B., C.M.M., D.Bak., D.O.O., D.P., D.A.R., D.S.Y.A., E.A.A., E.B., E.S.L., E.W., F.A.D., F.B., F.D., F.W., G.S., H.Ab., H.An., H.G., H.L., H.T., I.B.A., I.Ss., J.A.E., J.B., J.F., J.Gy., J.M.N., J.Y., K.H.Y., K.S., L.F., L.S., L.T., M.Ao., M.Al., M.G., M.T., M.V.T.P., M.R.W., N.D., N.Is., N.K., N.Ns., N.T., O.Cy., O.To., P.A., P.C.S., P.E.O., R.A.A.C., S.B., S.F.S., S.H.A., S.Kas., S.Ma., T.A., T.Mo., V.E., Y.Be. Funding acquisition: A.J.P., A.Re., A.v.G., B.Ko., C.N.A., C.A.K., C.B.P., C.W., D.C., D.J.B., D.J.N., F.L., G.A.A., G.G., G.P.M., H.C., J.E.S., J.-J.M.T., J.L., J.M.H., J.Nk., J.O., K.O.D., M-M.D., M.C., M.I.M., M.Mas., M.V.T.P., N.A., P.C.S., P.K., P.K.M., R.A.K., S.A.M., S.EI., S.Mo., S.v.d.W., T.d.O., W.P. Project administration: A.J.P., A.Re., A.F.V., A.v.G., B.Ko., C.W., D.J.B., D.J.N., E.W., F.Aj., F.T., G.A.A., G.P.M., G.S., G.T., H.C., J.C.O., J.-J.M.T., J.M.H., J.O., J.O.G., J.Y., K.O.D., M.C., M.K., M.Mar., M.P., M.V.T.P., M.R.W., N.R., O.To., P.C.S., P.K., P.K.M., R.A.K., S.A.M., S.EI., S.F.S., S.Gas., S.Mo., T.d.O. Supervision: A.J.P., A.Re., B.Ko., C.W., D.J.N., E.N.N., E.W., F.T., G.A.A., G.L.K., H.C., J.B., J.M.H., J.Nk., J.O., J.O.G., K.O.D., M.Al., M.C., M.I.M., M.Mar., M.M.N., M.S., N.Mu., N.R., P.C.S., P.K., P.K.M., R.A.K., S.EI., S.E.K., S.Gas., S.Me., S.Mo., S.Pa., T.d.O. Writing – original draft: A.K.Sa., A.-R.N.Z., B.Ko., D.P.M., E.W., F.T., G.L.K., H.T., J.B., J.-C.M., M.Al., M.A.B., M.C., M.G., M.Mi., N.Mu., R.L. Writing – review and editing: A.-R.N.Z., B.Ko., C.M.M., D.J.N., D.P.M., D.A.R., D.S.Y.A., D.T., E.K.L., E.L., E.S.L., E.W., H.T., J.E.S., J.G.S., L.d.O.M., M.A.B., M.C., M.e.H., P.K.Q., P.K.M., R.L., S.K.T., T.d.O., U.J.A. Author contributions are listed alphabetically. A full list of author abbreviations is included on the GitHub repository (https://github.com/krisp-kwazulu-natal/africa-covid19-genomics) (*23*). **Competing interests:** P.C.S. is a founder and shareholder of Sherlock Biosciences and is on the board and serves as shareholder of the Danaher Corporation. The authors declare no other conflicts of interest. **Data and materials availability:** All sequences that were used in the present study are listed in table S4 (accessible on the GitHub repository) along with their GISAID sequence IDs, dates of sampling, the originating and submitting laboratories, and main authors. All input files (e.g., alignments or XML files), all resulting output files, and scripts used in the study are shared publicly on GitHub (https://github.com/krisp-kwazulu-natal/africa-covid19-genomics) (*23*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to

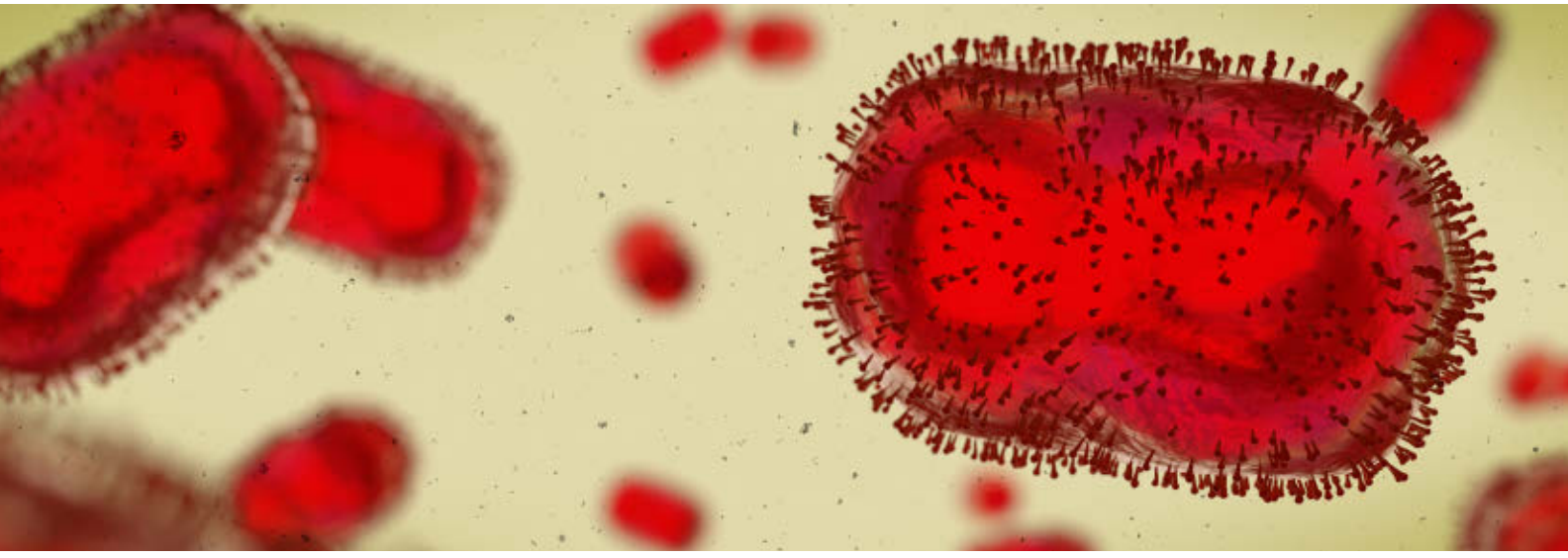
figures/photos/artwork or other content included in the article that is credited to a third party, obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj4336
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*24*–*38*)
MDAR Reproducibility Checklist

[1]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. [2]Centre for Epidemic Response and Innovation (CERI), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa. [3]Laboratorio de Flavivirus, Fundacao Oswaldo Cruz, Rio de Janeiro, Brazil. [4]Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [5]Department of Geography and GIS, University of Cincinnati, Cincinnati, OH, USA. [6]Institute of Infectious Diseases and Molecular Medicine, Department of Integrative Biomedical Sciences, Computational Biology Division, University of Cape Town, Cape Town, South Africa. [7]Division of Medical Virology, Wellcome Centre for Infectious Diseases in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. [8]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA. [9]Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA. [10]Cancer Biology Department, Virology and Immunology Unit, National Cancer Institute, Cairo University, Cairo 11796, Egypt. [11]Centre d'Infectiologie Charles Mérieux-Mali (CICM-Mali), Bamako, Mali. [12]Bacteriology and Virology Department Souro Sanou University Hospital, Bobo-Dioulasso, Burkina Faso. [13]MRCG at LSHTM Genomics Lab, Fajara, Gambia. [14]African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Osun State, Nigeria. [15]Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria. [16]Department of Medical Microbiology and Parasitology, Faculty of Basic Clinical Sciences, College of Health Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria. [17]CHU Habib Bourguiba, Laboratory of Microbiology, Faculty of Medicine of sFax, University of sFax, sFax, Tunisia. [18]Pathogen Sequencing Lab, Institut National de Recherche Biomédicale (INRB), Kinshasa, Democratic Republic of the Congo. [19]Université de Kinshasa (UNIKIN), Kinshasa, Democratic Republic of the Congo. [20]Genomics Research Program, Children's Cancer Hospital, Cairo, Egypt. [21]Institute of Pathogen Genomics, Africa Centres for Disease Control and Prevention (Africa CDC), Addis Ababa, Ethiopia. [22]Laboratory of Transmissible Diseases and Biological Active Substances (LR99ES27), Faculty of Pharmacy of Monastir, Monastir, Tunisia. [23]Laboratory of Microbiology, University Hospital of Monastir, Monastir, Tunisia. [24]Internal Medicine Department, Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [25]Irrua Specialist Teaching Hospital, Irrua, Nigeria. [26]Quadram Institute Bioscience, Norwich, UK. [27]World Health Organization, Africa Region, Brazzaville Congo. [28]Centre de Recherche et de Formation en Infectiologie de Guinée (CERFIG), Université de Conakry, Conakry, Guinea. [29]TransVIHMI, Montpellier University/IRD/INSERM, Montpellier, France. [30]Virology Department, Institut Pasteur de Dakar, Dakar, Senegal. [31]Mali-University Clinical Research Center (UCRC), Bamako, Mali. [32]Laboratory of Molecular and Cellular Screening Processes, Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [33]Broad Insitute of Harvard and MIT, Cambridge, MA, USA. [34]KEMRI-Wellcome Trust Research Programme/KEMRI-CGMR-C, Kilifi, Kenya. [35]National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service (NHLS), Johannesburg, South Africa. [36]The Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria [37]Institute of Virology, Charité – Universitätsmedizin, Berlin, Germany. [38]Medical Biotechnology Laboratory, Rabat Medical and Pharmacy School, Mohammed V University, Rabat, Morocco. [39]Department of Immunology, University of Maiduguri Teaching Hospital, P.M.B. 1414, Maiduguri, Nigeria. [40]MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda. [41]Division of Medical Virology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Cape Town, South Africa. [42]The Biotechnology Center of the University of Yaoundé I, Cameroon and CDC Foundation, Yaounde, Cameroon. [43]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. [44]Virology Unit, Institut Pasteur de Madagascar, Antananarivo, Madagascar. [45]University of Nebraska Medical Center (UNMC), Omaha, NE, USA. [46]Antibody Immunity Research Unit, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa. [47]CHU de Bouaké, Laboratoire/Unité de Diagnostic des Virus des Fièvres Hémorragiques et Virus Émergents, Bouaké, Côte d'Ivoire. [48]School of Public Health, Pwani University, Kilifi, Kenya. [49]Swiss Tropical and Public Health Institute, Basel, Switzerland. [50]West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana. [51]School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK. [52]Uganda Virus Research Institute, Entebbe, Uganda. [53]Division of Virology, National Health Laboratory Service and University of the Free State, Bloemfontein, South Africa. [54]Botswana Harvard AIDS Institute Partnership and Botswana Harvard HIV Reference Laboratory, Gaborone, Botswana. [55]University of Zambia, School of Veterinary Medicine, Department of Disease Control, Lusaka, Zambia. [56]African Society for Laboratory Medicine, Addis Ababa, Ethiopia. [57]Functional Genomic Platform/National Centre for Scientific and Technical Research (CNRST), Rabat, Morocco. [58]Rwanda National Reference Laboratory, Kigali, Rwanda. [59]Robert Koch-Institute, Berlin, Germany. [60]G5 Evolutionary Genomics of RNA Viruses, Institut Pasteur, Paris, France. [61]Research Unit of Autoimmune Diseases UR17DN02, Military Hospital of Tunis, University of Tunis El Manar, Tunis, Tunisia. [62]Department of Public Health, Ministry of Health, Ilorin, Kwara State, Nigeria. [63]Laboratory of Clinical Virology, Institut Pasteur de Tunis, Tunis, Tunisia. [64]Faculty of Pharmacy of Monastir, Monastir, Tunisia. [65]National Microbiology Reference Laboratory, Harare, Zimbabwe. [66]National Influenza Centre, Viral Respiratory Laboratory, Algiers, Algeria. [67]Medical Microbiology and Parasitology Department, College of Medicine, University of Ibadan, Ibadan, Nigeria. [68]Lincoln International Institute for Rural Health, University of Lincoln, Lincoln, UK. [69]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa. [70]Africa Health Research Institute, KwaZulu-Natal, Durban, South Africa. [71]Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya. [72]National Health Laboratory Service (NHLS), Tygerberg, Cape Town, South Africa. [73]Institution and Department, Ministry Of Health, COVID-19 Testing Laboratory, Mbabane, Kingdom of Eswatini. [74]Laboratory of Health Sciences and Technologies, High Institute of Health Sciences, Hassan 1st University, Settat, Morocco. [75]Department of Emerging Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan. [76]Central Public Health Laboratories (CPHL), Kampala, Uganda. [77]Faculty of Medicine Ain Shams Research institute (MASRI), Ain Shams University, Cairo, Egypt. [78]Charles Nicolle Hospital, Laboratory of Microbiology, National Influenza Center, 1006 Tunis, Tunisia. [79]Laboratory of Transmissible Diseases and Biological Active Substances (LR99ES27), Faculty of Pharmacy of Monastir, University of Monastir, Monastir, Tunisia. [80]Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel-Canada, Hadassah Medical School, The Hebrew University of Jerusalem, Jerusalem, Israel. [81]Namibia Institute of Pathology, Windhoek, Namibia. [82]Centre Interdisciplinaires de Recherches Medicales de Franceville (CIRMF), Franceville, Gabon. [83]Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA. [84]Urban Health Collaborative, Dornsife School of Public Health, Drexel University, Philadelphia, PA, USA. [85]UHAS COVID-19 Testing and Research Centre, University of Health and Allied Sciences, Ho, Ghana. [86]Rollins School of Public Health, Emory University, Atlanta, GA, USA. [87]Anoual Laboratory, Casablanca, Morocco. [88]Botswana Institute for Technology Research and Innovation, Gaborone, Botswana. [89]New York University Grossman School of Medicine, New York City, NY, USA. [90]Centre de Recherches Medicales de Lambarene (CERMEL), Lambarene, Gabon. [91]Virology/Molecular Biology Department, Central Health Laboratory, Ministry of Health and Wellness, Mauritius. [92]Center of Scientific Excellence for Influenza Viruses, National Research Centre (NRC), Cairo Egypt. [93]Ministry of Health and Wellness, Gaborone, Botswana. [94]Next Generation Sequencing Unit and Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein 9300, South Africa. [95]National Reference Laboratory Lesotho, Maseru, Lesotho. [96]Centre for Biotechnology Research and Development, Kenya Medical Research Institute, Nairobi, Kenya. [97]Laboratorio de Investigaciones de Baney, Baney, Equatorial Guinea. [98]Ifakara Health Institute, Dar-es-Salaam, Tanzania. [99]Nigeria Centre for Disease Control, Abuja, Nigeria. [100]Department of Medical Diagnostics, Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [101]Department of Medical Laboratory Science, Niger Delta University, Bayelsa State, Nigeria. [102]Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Cairo 12613, Egypt. [103]King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia. [104]Biological Prevention Department, Main Chemical Laboratories, Egypt Army, Cairo, Egypt. [105]PATH, Lusaka, Zambia. [106]Department of Biotechnology, High Institute of Biotechnology of Sidi Thabet, University of Manouba, BP-66, 2020 Ariana-Tunis, Tunisia. [107]Genomic Center for Human Pathologies (GENOPATH), Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco. [108]Rwanda National Joint Task Force COVID-19, Rwanda Biomedical Centre, Ministry of Health, Kigali, Rwanda. [109]School of Health Sciences, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [110]Instituto Nacional de Saude (INS), Maputo, Mozambique. [111]National Health Laboratory Service (NHLS), Cape Town, South Africa. [112]Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, Cape Town, South Africa. [113]Virology Laboratory, Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [114]Department of Epidemiology and Community Health, Faculty of Clinical Sciences, College of Health Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria. [115]Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [116]Mayotte Hospital Center, Mayotte, France. [117]Virology Service, Centre Pasteur of Cameroun, Yaounde, Cameroon. [118]Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [119]Laboratoire de Recherche et d'Analyses Médicales de la Gendarmerie Royale, Rabat, Morocco. [120]Clinical and Experimental Pharmacology Lab, LR16SP02, National Center of Pharmacovigilance, University of Tunis El Manar, Tunis, Tunisia. [121]CHU Hedi Chaker Sfax, Service de Pneumologie, Tunis, Tunisia. [122]Laboratoire de Recherche et d'Analyses Médicales de la Gendarmerie Royale, Rabat, Morocco. [123]Central Public Health Laboratories (CPHL), Cairo, Egypt. [124]Centre MURAZ, Ouagadougou, Burkina Faso. [125]National Institute of Public Health of Burkina Faso (INSP/BF), Ouagadougou, Burkina Faso. [126]National Reference Center for Respiratory Viruses, Molecular Genetics of RNA Viruses, UMR 3569 CNRS, University of Paris, Institut Pasteur, Paris, France. [127]Sub-Saharan African Network For TB/HIV Research Excellence (SANTHE), Durban, South Africa. [128]Coordenação Geral de Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil. [129]World Health Organization, WHO Lesotho, Maseru, Lesotho. [130]Med24 Medical Centre, Ruwa, Zimbabwe. [131]Division of Human Genetics, Department of Pathology, University of Cape Town, Cape Town, South Africa. [132]Center for Human Genetics, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [133]Laboratory of Human Genetics, GIGA Research Institute, Liège, Belgium. [134]National Health Laboratory, Gaborone, Botswana. [135]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK.[136]Harvard T.H. Chan School of Public Health, Boston, MA, USA. [137]Department of Global Health, University of Washington, Seattle, WA, USA. [138]Centre for Human Virology and Genomics, Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria. [139]School of Pathology, Faculty of Health Science, University of the Witwatersrand, Johannesburg, South Africa.

12 May 2021; accepted 3 September 2021
Published online 9 September 2021
10.1126/science.abj4336



NEWS

Monkeypox is a new global threat. African scientists know what the world is up against

Cases in West and Central Africa have been on the rise for decades

BY JON COHEN

As monkeypox stokes here-we-go-again fears in a pandemic-weary world, some researchers in Africa are having their own sense of déjà vu. Another neglected tropical disease of the poor gets attention only after it starts to infect people in wealthy countries. "It's as if your neighbor's house is burning and you just close your window and say it's fine," says Yap Boum, an epidemiologist in Cameroon who works with both the health ministry and Doctors Without Borders.

Now, the fire is spreading. The global outbreak of monkeypox, which causes smallpoxlike skin lesions but is not usually fatal, surfaced on 7 May in the United Kingdom. More than 700 suspected and confirmed cases had been reported as of 31 May, from every continent other than Antarctica. It is the largest ever outbreak outside of Africa and is concentrated among men who have sex with men, a phenomenon never seen before. Public health officials and scientists are scrambling to understand how the virus spreads and how to stop it—and they are paying new attention to Africa's long experience with the disease.

"We are interdependent," Boum notes. "What is happening in Africa will definitely impact what is happening in the West and vice versa."

Monkeypox is endemic in 10 countries in West and Central Africa, with dozens of cases this year in Cameroon, Nigeria, and the Central African Republic (CAR). The Democratic Republic of the Congo (DRC) has by far the highest burden, with 1284 cases in 2022 alone. Those numbers are almost certainly underestimates. In the DRC, infections most often happen in remote rural areas; in the CAR, armed conflict in several regions has limited surveillance.

The virus got its name after it was first identified in a colony of Asian monkeys in a Copenhagen, Denmark, laboratory in 1958, but it has only been isolated from a wild monkey—in Africa—once. It appears to be more common in squirrel, rat, and shrew species, occasionally spilling over into the human population, where it spreads mainly through close contact, but not through breathing. Isolating infected people typically helps outbreaks end quickly.

Cases have steadily increased in sub-Saharan Africa over the past 3 decades, driven largely by a medical triumph. The vaccine against smallpox, a far deadlier and more transmissible virus, also protects against monkeypox, but the world stopped using it in the 1970s, shortly before smallpox was declared eradicated. As a result, "There's a huge, huge number of people who are now susceptible to monkeypox," says Placide Mbala, a virologist who heads the genomics lab at the National Institute of Biomedical Research (INRB) in Kinshasa, DRC.

Mbala says demographic shifts have fueled the rise as well. "People are more and more moving to the forest to find food and to build houses, and this increases the contact between the wildlife and the population," he says. Studies in the CAR showed cases spike after villagers move into the forest during the rainy season to collect caterpillars that are sold for food. "When they stay in the bush they get in contact easily with the animal reservoir," says virologist Emmanuel Nakouné, scientific director at the Pasteur Institute of Bangui, which in 2018 launched a program named Afripox with French investigators to better understand and fight monkeypox.

Outbreaks outside Africa, including the current one, have all involved the West African strain, which kills about 1% of those it infects. The Congo Basin strain, found in the DRC and the CAR, is 10 times more lethal, yet despite the relatively high disease burden in the DRC, it has never left Africa. But it has never caused a serious outbreak in a Congolese city either, which underscores the isolation of the areas where it is endemic. "It's kind of a self-quarantine," Mbala says. "Those people don't move from DRC to other countries."

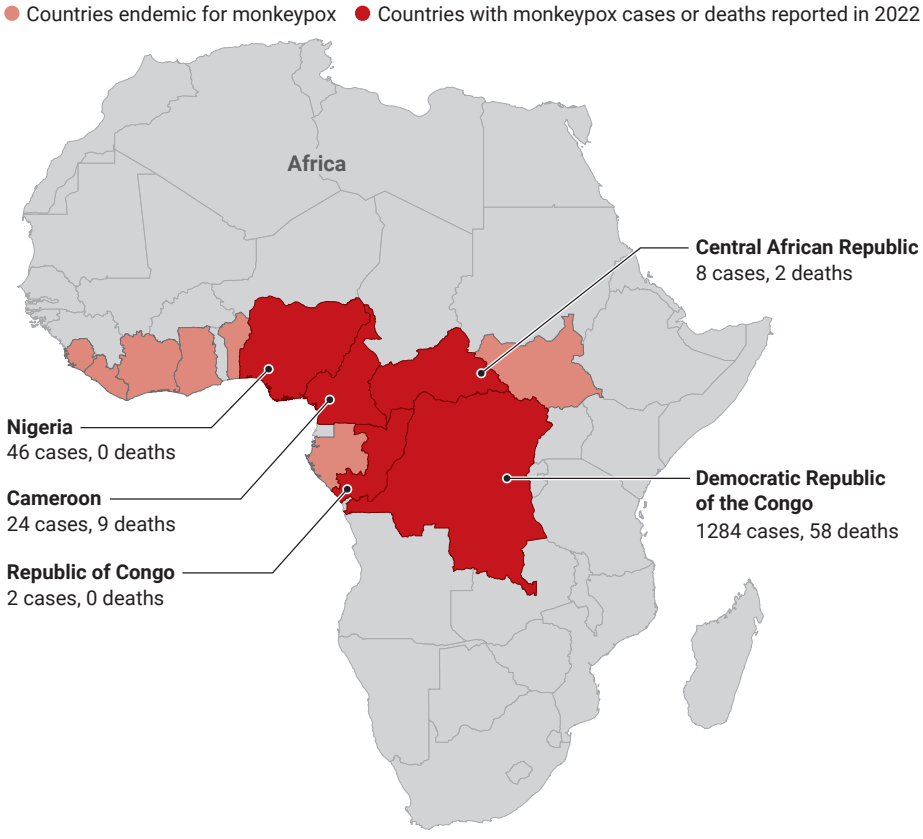
Just where the current outbreak started, and how long ago, is unclear. "It's a little bit like we've tuned into a new TV series and we don't know which episode we've landed on," says Anne Rimoin, an epidemiologist at the University of California, Los Angeles, who has worked on monkeypox in the DRC for 20 years. The first patient with an identified case traveled from Nigeria to the United Kingdom on 4 May, but does not appear to have infected anyone else. Two patients diagnosed later, one in the United States and the other in the United Arab Emirates, had recently traveled to Africa as well, and perhaps imported the virus separately. But none of the other cases identified in recent weeks has links to infected travelers or animals from endemic

IMAGE: © DOTTED YETI/SHUTTERSTOCK.COM

Originally published 1 June 2022

Spilling over

The monkeypox virus infects squirrel, rat, and shrew species in at least 10 countries in West and Central Africa and occasionally jumps into the human population. So far this year, five countries have reported human cases.



(GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) WORLD HEALTH ORGANIZATION

countries. Instead, many early cases were linked to transmission at gay festivals and saunas in Spain, Belgium, and Canada.

Some suspect the virus may have been imported from Nigeria, Africa's most populous country, which has good infrastructure connecting rural areas to large cities and two airports that are among the busiest in Africa. But this is "highly speculative," stresses Christian Happi, who runs Nigeria's African Centre of Excellence for Genomics of Infectious Diseases. Happi urges people in other countries "not to point fingers," but to collaborate.

Epidemiologist Ifedayo Adetifa, head of the Nigeria Centre for Disease Control, says the country receives undue attention because it does more surveillance than its neighbors and shares what it finds. "There's too much emphasis for whatever reasons in Western capitals and news media about trying to hold somebody responsible for a particular outbreak," he says. "We don't think those narratives are helpful." Adetifa says that although Nigeria has recently seen "an uptick in cases," he is confident it's not missing a large number of them. "We are literally rattling the bushes to see what comes out."

African countries' ability to deal with monkeypox was improving even before the current outbreak. The DRC has stepped up its surveillance across the vast country, which is key to isolating infected people and tracking the virus' moves. INRB and a lab in Goma can now diagnose samples using the polymerase chain reaction assay, and researchers ultimately hope to develop rapid tests for use in clinics nationwide. INRB and labs in Nigeria can also sequence the full genome of the virus, and Nigeria plans to make public genomes of several recent monkeypox isolates, Adetifa says. Those and other sequences from Africa could help researchers pinpoint the source of the international outbreak by building viral family trees.

For now, Africa lacks medicines to prevent and treat monkeypox. In the United Kingdom and the United States, high-risk contacts of cases are being offered a vaccine produced by Bavarian Nordic that was approved for monkeypox by the U.S. Food and Drug Administration in 2019, but it's not available anywhere in Africa. The U.S. Centers for Disease Control and Prevention and collaborators in the DRC are testing the vaccine in health care workers; the 2019 approval was based on animal studies.

In the CAR, 14 people with monkeypox have received an experimental drug, tecovirimat, as part of a trial launched by the University of Oxford in July 2021. "We've had very good results," says Nakouné, who says he expects the data to be published within the next few weeks. The drug's manufacturer, SIGA, has pledged to provide up to 500 treatment courses to the country.

Although the international outbreak has—again—highlighted global health inequities, it has also brought much-needed attention to the smoldering disease in Africa. "It's been really hard to get the resources to do the kind of background work that really needs to be done and that isn't hair-on-fire, in the context of an emergency," Rimoin says. "We cannot keep hitting the snooze button. Now, the stakes are really high."

CORONAVIRUS

The puzzle of the COVID-19 pandemic in Africa

More data are needed to understand the determinants of the COVID-19 pandemic across Africa

By Justin M. Maeda and John N. Nkengasong

The COVID-19 pandemic has been puzzling to many public health experts because Africa has reported far fewer cases and deaths from COVID-19 than predicted. As of 22 November 2020, the continent of Africa, comprising 1.3 billion people, had recorded 2,070,953 cases of COVID-19 and 49,728 deaths (*1*), representing ~3.6% of total global cases (*2*, *3*). Because of the continent's overstrained and weak health systems, inadequate financing of health care, paucity in human resources, and challenges posed by existing endemic diseases—including HIV, tuberculosis, and malaria—earlier predictions suggested that up to 70 million Africans may be infected with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) by June, with more than 3 million deaths (*4*). On page 79 of this issue, Uyoga *et al.* (*5*) report a serosurvey study (measuring the occurrence of SARS-CoV-2 antibodies) of blood donors in Kenya that suggested that the incidence of SARS-CoV-2 infection is much higher than expected from case numbers.

Using blood donor samples as a proxy, Uyoga *et al.* estimated that SARS-CoV-2 infections occurred in 5.5% of the population in Kisumu, 7.3% in Nairobi, and 8.0% in Mombasa, with an overall average of 4.3%. This translates to ~2.2 million total possible infections compared with the reported 77,585 infections in the country as of 23 November 2020 (*1*, *3*). Similarly, in October 2020, Mozambique reported less than 3000 confirmed cases of COVID-19; however, serosurveys found that 5% of households in the city of Nampula and 2.5% of households in the city of Pemba had been exposed to the virus (*6*). This suggests that there may be more infections than recorded.

There are several factors that may influence the trajectory of the COVID-19 pandemic in Africa. These include limited testing (which limits detection and isolation, and thus public health measures), a much younger population (and thus fewer severe cases and deaths), climatic differences (which could affect transmission), preexisting immunity, genetic factors, early implementation of public health measures, and timely leadership. Two key aspects that may contribute to our understanding of the pandemic puzzle in Africa include scaling up of testing and use of serosurveys.

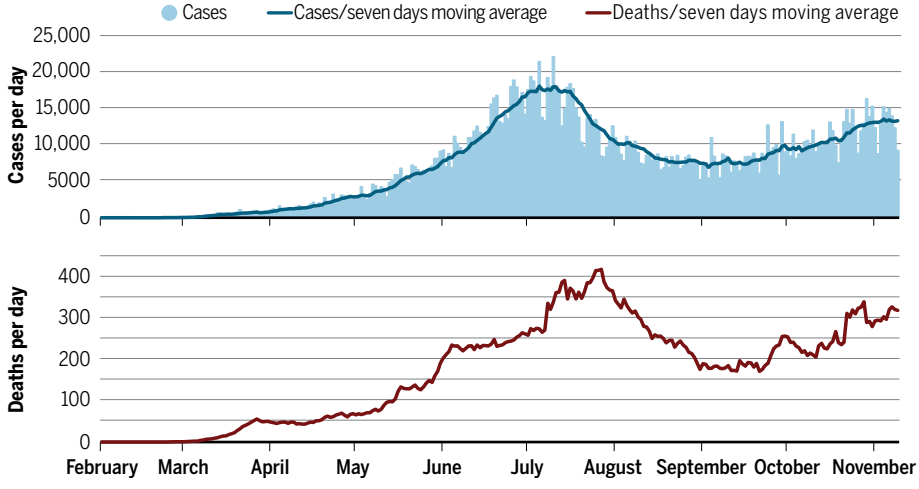
One way to unravel the puzzle of SARS-CoV-2 spread is to understand how testing and reporting of cases has occurred. On 14 February 2020, the first cases of COVID-19 were reported in Africa, and by 17 April 2020, the continent had conducted an estimated 330,419 SARS-CoV-2 tests; that is 0.03% of the entire continent's population. In an effort to scale up testing, the Africa Centres for Disease Control and Prevention (Africa CDC) launched the Partnership to Accelerate COVID-19 Testing (PACT) in April 2020. Because of the PACT initiative, testing was scaled up rapidly from ~600,000 per month in April to ~3.5 million per month in November 2020, an increase of ~5.5-fold (*7*), with 39 of 55 (71%) countries reporting more than 10 tests conducted for every case identified, as recommended by the World Health Organization (WHO) (*8*). Testing capacity has varied over time, with positivity rate fluctuating between 5 and 15% regardless of the increased testing boost brought by the introduction of PACT.

Therefore, it is clear that testing has been challenging (*9*, *10*), which limits our understanding of the full extent of the spread of SARS-CoV-2 infection in Africa. As such, serosurveys are critical because they can provide data on SARS-CoV-2 infection trends, effects of interventions, demographic characterization, and vaccine effects. Such surveys can also inform on planning for vaccine deployment by providing data to guide prioritization between different populations. They can also aid understanding of the drivers of infection through linking current or previous infection with epidemiological and demographic data. Currently, the continent is facing a challenging phase of the pandemic with an observed second wave of cases (see the figure). More people need to be tested in different localities, including repeated testing over time, so that the patterns and risk factors of viral spread can be understood.

Several serosurveys have been conducted in Africa. The studies differ in methodological approach used: simple random sampling, use of existing sentinel sites, and targeted population (specific subnational unit, pregnant women, blood donors, and people living with HIV). The types of laboratory tests used (rapid tests and enzyme-linked immunosorbent assays) also differed between studies so as to unveil the drivers of infection and disease spread. Given the limited ability to conduct field surveys (the preferred method)

COVID-19 cases and deaths in Africa

The trend of daily reported cases of COVID-19 for the African continent, February to November 2020, shows the first peak of cases occurred July to August (mostly attributed to the Southern African Region) followed by a second peak, which started in October (mostly attributed to the Northern Region).



Africa Centres for Disease Control and Prevention, Addis Ababa, Ethiopia. Email: justinm@africa-union.org; nkengasongJ@africa-union.org

Originally published 1 January 2021

owing to travel restrictions, Uyoga *et al.* investigated blood donors to reveal the pandemic puzzle in Kenya. From these surveys in Africa, seroprevalence of SARS-CoV-2–specific antibodies have ranged from 2.2 to 39% of the population in different settings and countries. However, none of the studies have used a national representative sample.

To ensure a harmonized and standardized method of conducting serosurveys in Africa, the Africa CDC is supporting multinational population-based, age, and gender stratified serosurveys that use standardized protocol and data collection tools (*11*). The protocol is built on a simplistic model, using point-of-care rapid test for antibody detection of current and previous infection, to ensure feasibility and simplicity while maintaining study quality and credibility of the evidence generated. A similar approach has been applied to national representative cohorts in Brazil and Spain (*12*, *13*).

Across Africa, policy makers are faced with the dilemma of striking a balance between limiting transmission and protecting economies, businesses, and livelihoods. This has created a demand for quality and comprehensive data. Serosurveys could therefore complement existing response strategies. Such surveys should adhere to the following principles: a national representative sample through well-designed sampling strategies that ensure inclusivity of all possible strata within the country; simplicity to guarantee feasibility and quick delivery; optimization of resources for implementation (human, material, and financial) to safeguard the already constrained resources for response; complementarity to already existing surveillance and response data; and the ability to longitudinally track the same aspect of data and information over time to inform adaptive strategies.

Timely leadership and coordination may be a second aspect that explains the pandemic pattern in Africa. The continent reacted in a timely and collective manner once the first cases of SARS-CoV-2 were reported in Egypt on 14 February 2020. Following that, on 22 February 2020, the Africa CDC convened an emergency meeting of all ministers of health at the headquarters of the African Union Commission in Addis Ababa, Ethiopia. The ministers adopted a joint continental strategy that had three goals: limit transmission, limit deaths, and limit social and economic harms and impacts on other endemic diseases, underpinned by the need to coordinate, cooperate, collaborate, and communicate efforts across Africa. In addition, the Africa Taskforce on Coronavirus (AFTCOR) was established to help implement the strategy and was endorsed by the Bureau of the Heads of State and Governments of the African Union, a validation at the highest level of the continent. This approach helped blunt the early spread of COVID-19.

> ***“Across Africa, policy makers are faced with the dilemma of striking a balance between limiting transmission and protecting economies, businesses, and livelihoods.”***

Therefore, in March 2020, when several countries in Africa began reporting imported cases of COVID-19, there was clarity on the course of action to take. For example, as part of the AFTCOR, the Africa CDC rapidly supported member states to establish diagnostics capacity and expanded testing capacity from two countries in February to more than 43 by the end of March, through competency-based training of member countries at reference centers in Dakar, Senegal, and Johannesburg, South Africa. The coordinated approach ensured harmony in response strategies. For example, the establishment of the African Medical Supply Platform helped to streamline the procurement of response commodities.

The puzzle of the COVID-19 pandemic in Africa can partly be explained by decisive measures taken early to prepare the continent. However, more data are needed to complement what is routinely collected through surveillance and response to understand the different pieces of the puzzle that contribute to the pattern of the pandemic in Africa. Serosurveys and the use of genomic epidemiology can help to better understand disease spread. Further understanding of factors that influence viral pathogenesis and clinical spectrum of disease, and the impacts on endemic infections (HIV, tuberculosis, and malaria), are needed. Efforts to understand attitudes to COVID-19 vaccines are also a priority. ■

REFERENCES AND NOTES

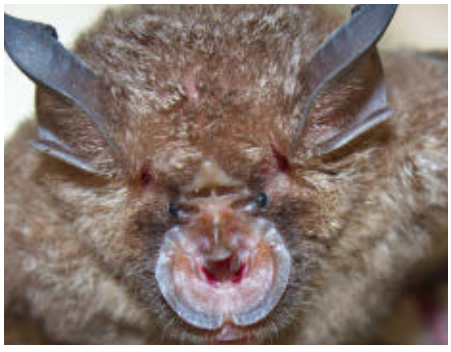
1. Africa Centres for Disease Control and Prevention, Latest updates on the COVID-19 crisis in Africa (2020); https://africacdc.org/covid-19.
2. World Health Organization, WHO coronavirus disease (COVID-19) dashboard (2020); https://covid19.who.int.
3. United Nations Department of Economy and Social Affairs, World population prospect 2019 (2019); https://population.un.org/wpp.
4. P. G. Walker, C. Whittaker, O. Watson, Report 12 - The global impact of COVID-19 and strategies for mitigation and suppression (2020); https://bit.ly/37SEG8t.
5. S. Uyoga *et al.*, *Science* **371**, 79 (2020).
6. A. Frey, Mozambique: 3.79 per cent of Maputo residents exposed to coronavirus (Club of Mozambique, 2020); https://bit.ly/37Si9J1.
7. C. D. C. Africa, Outbreak brief 45: Coronavirus disease 2019 (COVID-19) pandemic (2020); https://bit.ly/3qL0KuE.
8. Our World in Data–University of Oxford, Test conducted per new case of covid-19; https://bit.ly/2Kbzd4y.
9. J. Nkengasong, *Nature* **580**, 565 (2020).
10. J. N. Nkengasong, N. Ndembi, A. Tshangela, T. Raji, *Nature* **586**, 197 (2020).
11. C. D. C. Africa, Generic protocol for a population-based, age- and gender- stratified sero-survey study for SARS-CoV-2 (2020); https://bit.ly/3n5QuuL.
12. P. C. Hallal *et al.*, *Lancet Glob. Health* **8**, e1390 (2020).
13. M. Pollán *et al.*, *Lancet* **396**, 535 (2020).

10.1126/science.abf8832

SARS-like viruses may jump from animals to people hundreds of thousands of times a year

NEWS

Study pinpoints Asian regions that could spark the next coronavirus pandemic



By Kai Kupferschmidt

Only two new coronaviruses have spread globally the past 2 decades: SARS-CoV, which caused an outbreak of severe acute respiratory syndrome (SARS) in 2003, and SARS-CoV-2, the virus that causes COVID-19. But that may just be the tip of the iceberg of undetected infections with related viruses emerging from bats, a new paper claims. In a preprint published yesterday[1] researchers estimate that an average of 400,000 people are likely infected with SARS-related coronaviruses every year, in spillovers that never grow into detectable outbreaks.

Although that number comes with big caveats, "It should be eye-opening to the entire scientific community that we don't know very much about the frequency of zoonotic spillover," says virologist Angela Rasmussen of the University of Saskatchewan, who was not involved in the work. That needs to change, she says, "because otherwise we grossly underestimate it."

The researchers, including Peter Daszak from the EcoHealth Alliance and Linfa Wang from Duke-NUS Medical School in Singapore, created a detailed map of the habitats of 23 bat species known to harbor SARS-related coronaviruses, the group to which SARS-CoV and SARS-CoV-2 belong, and then overlaid it with data on where humans live to create a map of potential infection hot spots. They found that close to 500 million people live in areas where spillovers can occur, including northern India, Nepal, Myanmar, and most of Southeast Asia. The risk is highest in southern China, Vietnam, Cambodia, and on Java and other islands in Indonesia (see map, below).

"This is a definitive analysis of where on the planet the next SARS- or COVID-like virus is most likely to emerge," Daszak says. The maps could guide efforts to reduce the likelihood of spillover by changing behaviors in high-risk communities and targeting surveillance to detect new outbreaks earlier, he says. Daszak, a vocal advocate of the hypothesis that SARS-CoV-2 came from the wild instead of a research lab, says the maps could also guide efforts to find the virus' natural origin. (Several studies are underway or being planned to look for SARS-CoV-2 and its relatives in *Rhinolophus* [horseshoe] bats and other animals.)

But the researchers went one step further. Small surveys done before COVID-19 erupted have suggested some people in Southeast Asia harbor antibodies against SARS-related coronaviruses. Combining those data with data on how often people encounter bats and how long antibodies remain in the blood, the researchers calculated that some 400,000 undetected human infections with these viruses occur each year across the region.

Daszak says interactions with bats are much more common than people think: "Just living there means you're exposed: People are sheltering in caves, they're digging guano out of caves, they're hunting and eating bats." The paper does not even address how many people work in the wildlife trade and may be infected indirectly when a bat virus infects another animal first, he says.

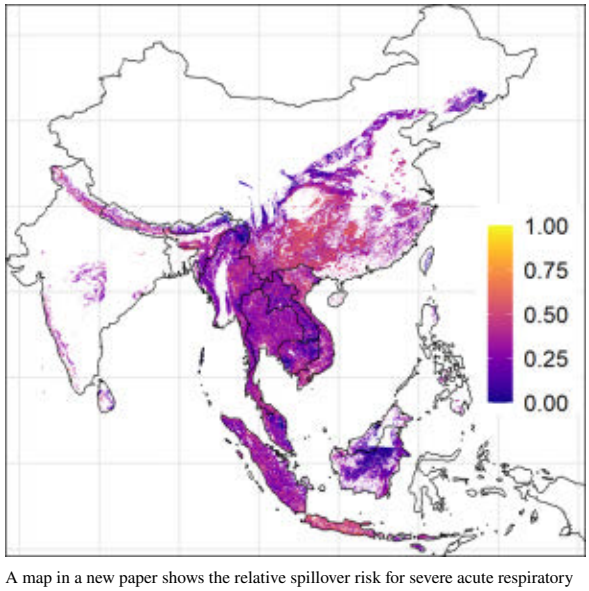
Although 400,000 infections annually sounds like a lot, Rasmussen says, "in a region with likely hundreds of millions of bats and nearly half a billion people it isn't that many." The confidence interval stretches from one to more than 35 million hidden infections per year, however—big enough to "fly the entire population of Rhinolophid bats through," Rasmussen quips. Models are only as good as the data that are fed into them, says Vincent Munster, a virologist at the U.S. National Institute of Allergy and Infectious Diseases who studies coronaviruses. The data on antibodies only include a few thousand people, he notes, and the assays used to test for antibodies can easily lead to false positives.

"I think if the seroprevalence estimate is way off, the whole thing collapses," says David Fisman, an epidemiologist at the University of Toronto, who calls the modeling "shaky." The high number of hidden infections "doesn't ring true," Fisman says, because you would expect regular spillovers to be recognized, as they are for rabies and the Nipah virus.

But Rasmussen says many infections could remain hidden if they are short-lived and don't lead to onward transmission because the viruses are not well adapted to humans. They might not infect enough cells—or cells of the right type—to be transmitted to another person, or they might not be able to escape humans' immune defenses. In cases when the virus does spread, sheer chance may keep it confined to a small, isolated community.

"A lot of the viruses are probably unable to be transmitted from one person to another, but I have very little doubt that there have been illnesses due to these viruses that get misdiagnosed or never diagnosed," Daszak says. "A rural farmer in Myanmar is hardly likely to go to the clinic because they've got a bit of a cough."

The work is part of a nascent effort to try to understand the risk factors for viral spillover from animals into humans, Munster says. Already, one message is clear, he says: "I think for virtually any zoonotic pathogen from wildlife, spillover is more frequent than previously recognized."



A map in a new paper shows the relative spillover risk for severe acute respiratory syndrome–related coronaviruses. China and countries in Southeast Asia are potential hot spots for human infections.

PHOTO: © BINTURONG-TONOSCARPE/SHUTTERSTOCK.COM; GRAPHIC: C. A. SÁNCHEZ *ET AL.*, MEDRXIV (2021) 10.1101/2021.09.09.21263359

[1] https://www.medrxiv.org/content/10.1101/2021.09.09.21263359v1

INFECTIOUS DISEASES

The Global Virome Project

Expanded viral discovery can improve mitigation

By Dennis Carroll, Peter Daszak, Nathan D. Wolfe, George F. Gao, Carlos M. Morel, Subhash Morzaria, Ariel Pablos-Méndez, Oyewale Tomori, Jonna A. K. Mazet

Outbreaks of novel and deadly viruses highlight global vulnerability to emerging diseases, with many having massive health and economic impacts. Our adaptive toolkit—based largely on vaccines and therapeutics—is often ineffective because countermeasure development can be outpaced by the speed of novel viral emergence and spread. Following each outbreak, the public health community bemoans a lack of prescience, but after decades of reacting to each event with little focus on mitigation, we remain only marginally better protected against the next epidemic. Our ability to mitigate disease emergence is undermined by our poor understanding of the diversity and ecology of viral threats, and of the drivers of their emergence. We describe a Global Virome Project (GVP) aimed to launch in 2018 that will help identify the bulk of this viral threat and provide timely data for public health interventions against future pandemics.

Nearly all recent pandemics have a viral etiology with animal origins, and with their intrinsic capacity for interspecies transmission, viral zoonoses are prime candidates for causing the next great pandemic (*1*, *2*). However, if these viruses are our enemy, we do not yet know our enemy very well. Around 263 viruses from 25 viral families are known to infect humans (*3*) (see the figure), and given the rate of discovery following identification of the first human virus (yellow fever virus in 1901), it is likely many more will emerge in the future (*4*). We estimate, from analysis of recent viral discovery data (*5*), that ~1.67 million yet-to-be-discovered viral species from key zoonotic viral families exist in mammal and bird hosts—the most important reservoirs for viral zoonoses (supplementary text).

By analyzing all known viral-host relationships (*3*, *6*), the history of viral zoonoses (*7*), and patterns of viral emergence (*1*), we can reasonably expect that between 631,000 and 827,000 of these unknown viruses have zoonotic potential (supplementary text). We have no readily available technological countermeasures to these as-yet-undiscovered viruses. Furthermore, the rate of zoonotic viral spillover into people is accelerating, mirroring the expansion of our global footprint and travel networks (*1*, *8*), leading to a nonlinear rise in pandemic risk and an exponential growth in their economic impacts (*8*).

PROMISING PILOT, CHALLENGING SCALE

Since 2009, the U.S. Agency for International Development (USAID) has conducted a large-scale pilot project, spanning more than 35 countries over 8 years at a cost of around $170 million, to evaluate the feasibility of preemptively mitigating pandemic threats. Other previous studies had begun to conduct targeted viral discovery in wildlife (*9*), and develop mitigation strategies for the emergence of avian flu, for example. However, the USAID Emerging Pandemic Threats (EPT) PREDICT project is the first global-scale coordinated program designed to conduct viral discovery in wildlife reservoir hosts, and characterize ecological and socioeconomic factors that drive their risk of spillover, to mitigate their emergence in people (*10*).

Working with local partners and governments, wildlife and domestic animals and at-risk human populations in geographic hotspots of disease emergence (*1*) are sampled, and viral discovery conducted. A strategy to identify which novel viruses are most at risk of spillover has been developed (*11*), and further work is conducted on these to characterize them prior to, or in the early stages of, spillover. Metadata on the ecology of wildlife–livestock–human transmission interfaces, and on human behavioral patterns in communities, are concurrently analyzed so that strategies to reduce spillover can be developed (supplementary text). To date, EPT PREDICT has discovered more than 1000 viruses from viral families that contain zoonoses, including viruses involved in recent outbreaks (*12*), and others of ongoing public health concern (*13*). The focus of EPT PREDICT on capacity building, infrastructure support, training, and epidemiological analysis differs substantially from the GVP's emphasis on large-scale sampling and viral discovery. However, to discover the bulk of the projected remaining 1.67 million unknown viruses in animal reservoirs and characterize the majority of 631,000 to 827,000 viruses of highest zoonotic potential requires overcoming some challenges of scale.

The first challenge is cost. To estimate this, we analyzed data on field sampling and laboratory expenditures for viral discovery from (*5*, *10*), and estimates of unknown viral diversity in mammalian and avian hosts (supplementary text). We estimate that discovery of all viral threats and characterization of their risk for spillover, using currently available technologies and protocols, would be extremely costly at over $7 billion (supplementary text). However, previous work shows that viral discovery rates are vastly higher in the early stages of a sampling program, and that discovering the last few, rare, viruses is extremely costly and time-consuming owing to the number of samples required to find them (*5*) (supplementary text). We used data on rates of viral discovery (*5*) to estimate that the substantial majority of the viral diversity from our target zoonotic reservoirs could be discovered, characterized, and assessed for viral ecology within a 10-year time frame for ~$1.2 billion (16% of total costs for 71% of the virome, considering some fixed costs) (fig. S1). Those viruses remaining undiscovered will, by the nature of sampling bias toward more common host species, represent the rarest viruses with least opportunity for spillover, and therefore reduced public health risk. Their discovery would require exponentially greater sampling effort and funding that could be better spent on countermeasures for the more likely threats (supplementary text).

Stakeholders from Asia, Africa, the Americas, and Europe, spanning industry, academia, intergovernmental agencies, nongovernmental organizations (NGOs), and the private sector, began meeting in 2016 to design a framework for the governance, management, technical operation, and scope of the GVP. Key efforts include developing finance streams; establishing a transparent, equitable implementation strategy; designing data- and sample-sharing protocols; developing laboratory and metadata platforms; targeting of host taxa and sampling sites; analyzing return on investment; forming collaborative field and laboratory networks; developing risk characterization frameworks for viruses discovered; designing a strategy to assess and mitigate risk behaviors that facilitate viral emergence; and planning in-country capacity building for sustainable threat mitigation. Funding has been identified to support an initial administrative hub, and fieldwork is planned to begin in the first two countries, China and Thailand, during 2018.

With outputs intended to serve the global public good, the GVP is developing a transparent and equitable strategy to share data, viral samples, and their likely products, including benefits derived from future development of medical countermeasures. These build on the Nagoya Protocol to the Convention on Biological Diversity and the Pandemic Influenza Preparedness Framework, negotiated by the World Health Organization (WHO). The international collaborative nature and global ownership of the project should help leverage funding from diverse international donors, including government agencies focused on national virome proj-

The list of author affiliations is provided in the supplementary materials. Email: daszak@ecohealthalliance.org

Originally published 23 February 2018

ects or on international development projects in other countries, and private-sector philanthropic donors focused on technology and big science.

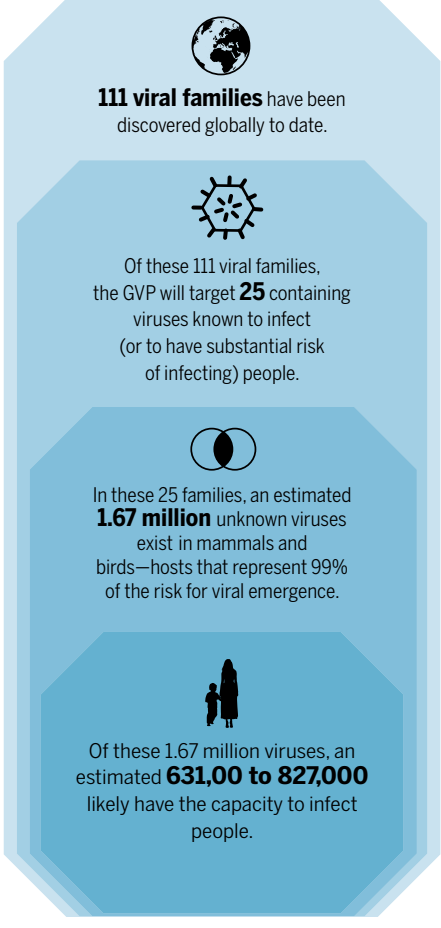
The diversity of tasks required to conduct the GVP should reduce the potential for it to divert funds from current public health programs. For example, discrete work streams on targeted sampling of wildlife, on bioinformatics, and on behavioral risk analysis fall within the focus of current scientific research programs in a range of donor agencies. Governments and corporations with specific remits and geographic responsibilities have been approached to finance subprojects relevant to their sectors (e.g., capacity development, surveillance of specific taxa, geographically focused activities, medical countermeasure development, training, surveillance, and technological platforms). In addition, leaders in China and a number of countries have begun developing national virome projects as part of the GVP, leveraging current research funding to include GVP sites.

Technological challenges include safe field sampling in remote locations and cost-effective laboratory platforms that can be standardized in low-income settings. To achieve these goals, existing national, regional, and international networks will need to be enhanced and expanded within standardized sampling and testing frameworks. Existing networks of field biologists from environment ministries, academic institutions, and conservation and health NGOs may assist in surveillance. National science and technology agencies, regional One Health platforms, transboundary disease surveillance networks, Institut Pasteur laboratories, WHO, United Nations Food and Agricultural Organization, and the World Organization for Animal Health collaborating, and reference centers and viral discovery laboratories, including USAID EPT PREDICT, are currently involved in planning these activities around a decade-long sampling and testing time frame. A monitoring and evaluation strategy is being developed based on analysis of viral discovery rates against predicted viral diversity, to identify when to halt surveillance and testing as the GVP progresses. Stakeholders will also tackle the challenge of how to decide when enough potentially dangerous viruses have been discovered in a host species or region to call for action to reduce underlying drivers of emergence (e.g., hunting and trading of a wildlife reservoir). Laboratory platforms developed by USAID EPT PREDICT have proven capacity to identify novel viruses and are relatively inexpensive and reliable, being based on polymerase chain reaction using degenerate primers that target a range of viral families of known zoonotic potential. However, scaling up to a full global virome project will require discovery of three orders of magnitude more viruses in a similar time frame. Technological solutions will be needed to increase the speed and efficiency, and reduce the cost, of sequence generation. These will likely include next-generation sequencing and other unbiased approaches to identify evolutionarily distinct viral clades.

A key challenge is how to assess the potential for novel viruses to infect people or become pandemic (*14*). The EPT PREDICT project (*11*) and others (*2*, *6*) have developed preliminary zoonotic risk characterization frameworks based on viral and host traits and the ecological and demographic characteristics of the sampling site. These approaches will be used in the GVP to triage novel viruses for further characterization to assess their zoonotic capacity (supplementary text). In vitro receptor binding analyses coupled with in vivo models have proven useful in this capacity for some viral families [e.g., coronaviruses (*13*)]. Although this is not yet feasible for all potentially zoonotic viral clades, applying these techniques to a larger viral data set as the GVP progresses will allow validation of risk frameworks and may increase our capacity to predict zoonotic potential. However, advancing these goals will require new collaboration among lab virologists, epidemiologists, and modelers, innovative approaches to field-testing the boundaries of virus-host relationships, and support across agencies that often fund separate virology, public health, evolutionary biology, and biodiversity modeling initiatives.

GVP targeting strategy

The project will capitalize on economies of scale in viral testing, systematically sampling mammals and birds to identify currently unknown, potentially zoonotic viruses that they carry.


111 viral families have been discovered globally to date.
Of these 111 viral families, the GVP will target **25** containing viruses known to infect (or to have substantial risk of infecting) people.
In these 25 families, an estimated **1.67 million** unknown viruses exist in mammals and birds—hosts that represent 99% of the risk for viral emergence.
Of these 1.67 million viruses, an estimated **631,00 to 827,000** likely have the capacity to infect people.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

INVESTMENTS, RETURNS

The cost of the GVP represents a sizable investment and, even if a large number of potential zoonoses are discovered, only a minority is likely to have the potential to cause large-scale outbreaks and mortality in people (*1*, *2*, *7*). However, given the high cost of single epidemic events, data produced by the GVP may provide substantial return on investment by enhancing diagnostic capacity in the early stages of a new disease outbreak or by rapidly identifying spillover hosts, for example. Recent analysis of the exponentially rising economic damages from increasing rates of zoonotic disease emergence suggests that strategies to mitigate pandemics would provide a 10:1 return on investment (*1*, *8*). Even small reductions in the estimated costs of a future influenza pandemic (hundreds of billions of dollars) or of the previous SARS (severe acute respiratory syndrome) epidemic ($10 billion to $30 billion) could be substantial. The goal of the GVP is to improve efficiency in the face of these increasing viral spillover rates by enhancing (not replacing) current pandemic surveillance, prevention, and control strategies. If we were to invest only in surveillance for known pathogens (our current business-as-usual strategy), our calculations suggest we would protect ourselves against less than 0.1% of those viruses that could conceivably infect people, even using the lower bounds of our uncertainty for our viral estimates (i.e., 263 viruses known from humans out of 263,824 unknown potential zoonoses; supplementary text).

The potential benefits of the GVP may be enhanced to maximize public health benefits (supplementary text) by (i) optimizing sampling to target species most likely to harbor "missing zoonoses" (*6*), or to target emerging disease hotspot regions most likely to propagate major disease outbreaks (*1*); (ii) using human and livestock syndromic surveillance to identify regions for wildlife sampling proximal to repeated outbreaks of severe influenza-like-illnesses, fevers of unknown origin, encephalitides, livestock "abortion storms," and other potential emerging disease events; (iii) initially targeting RNA viruses, which caused 94% of the zoonoses documented from 1990 to 2010; and (iv) fostering

economies of scale and adoption of technological innovation as the GVP ramps up. This includes use of laboratories that can facilitate regional sample processing, development of centralized bioinformatics platforms, and improved logistics for sample collection and transport. We also expect the cost of testing and sequencing to decrease as technology is enhanced, much as the development of next-generation sequencing reduced genetic sequencing costs by up to four orders of magnitude in a decade.

The accelerated pace of viral discovery under the GVP will make the virological, phylogenetic, and modeling approaches used in pandemic preparedness more data-rich, and likely more effective. For example, having the sequence data for thousands, rather than a few, viruses from a single family could extend vaccine, therapeutic, or drug development to a wider range of targets, leading to broad-spectrum vaccines and other countermeasures. Identification of novel viruses may be useful to programs like the Coalition for Epidemic Preparedness Innovations (CEPI) in assessing the breadth of action of candidate vaccines and therapeutics, and in expanding their efficacy. More broad-scale prevention approaches could provide immediate return on investment prior to vaccine and countermeasure development, which would require substantial investment and time. For example, metadata on viral reservoir host identity, geography, seasonality, proximity to people, and drivers of emergence will refine our mechanistic understanding of spillover and enhance published models of emerging infectious diseases risk (*1*, *6*). Identification of novel viruses in hunted, traded, or farmed wildlife species could be used to enhance biosecurity in markets and farming systems, reducing public health risk, increasing food security, and assisting in conservation of hunted species. The presence of hosts harboring high-risk novel viruses in proximity to human populations may allow targeted follow-up to examine evidence of spillover and design intervention strategies (supplementary text). Ultimately, the benefits of the GVP may include enhancing our understanding of viral biology, such as drivers of competition or cooperation among viruses within hosts, genomic underpinnings of host-virus coevolution, processes underlying deep evolution of viral clades, and the identification of novel viral groups (*15*).

The regions targeted by the GVP are largely highly biodiverse, rapidly developing countries in the tropics, which often have low capacity to deal with public health crises (*1*). The expanded laboratory capacity, field sampling, and data generation intrinsic to the GVP goals will therefore improve capacity to detect, diagnose, and discover viruses in vulnerable populations within regions most critical to preventing future pandemics. This enhanced capacity may also help improve diagnosis and control for endemic diseases, as well as the portion of the virome that remains undiscovered.

The Human Genome Project in the 1980s catalyzed technological innovation that dramatically shortened the time and cost for its completion, and ushered in the era of personalized genomics and precision medicine. The GVP will likely accelerate development of pathogen discovery technology, diagnostic tests, and science-based mitigation strategies, which may also provide unexpected benefits. Like the Human Genome Project, the GVP will provide a wealth of publicly accessible data, potentially leading to discoveries that are hard to anticipate, perhaps viruses that cause cancers and chronic physiological, mental health, or behavioral disorders. It will provide orders-of-magnitude more information about future threats to global health and biosecurity, improve our ability to identify vulnerable populations, and enable us to more precisely target mitigation and control measures to foster an era of global pandemic prevention.

REFERENCES AND NOTES

1. K. E. Jones *et al.*, *Nature* **451**, 990 (2008).
2. J. L. Geoghegan, A. M. Senior, F. Di Giallonardo, E. C. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4170 (2016).
3. A. M. King, M. J. Adams, E. B. Carstens, E. J. Lefkowitz, *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses* (Elsevier, Philadelphia, 2012).
4. M. Woolhouse, F. Scott, Z. Hudson, R. Howey, M. Chase-Topping, *Philos. Trans. R. Soc. Lond. B* Biol. Sci. **367**, 2864 (2012).
5. S. J. Anthony *et al.*, *mBio* **4**, e00598-13 (2013).
6. K. J. Olival *et al.*, *Nature* **546**, 646 (2017).
7. N. D. Wolfe, C. P. Dunavan, J. Diamond, *Nature* **447**, 279 (2007).
8. J. Pike, T. L. Bogich, S. Elwood, D. C. Finnoff, P. Daszak, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18519 (2014).
9. J. F. Drexler *et al.*, *Nat. Commun.* **3**, 796 (2012).
10. data.predict.global
11. S. S. Morse *et al.*, *Lancet* **380**, 1956 (2012).
12. G. Grard *et al.*, *PLOS Pathogens* **8**, e1002924 (2012).
13. X.-Y. Ge *et al.*, *Nature* **503**, 535 (2013).
14. J. L. Geoghegan, E. C. Holmes, *Open Biol.* **7**, 170189 (2017).
15. C. X. Li *et al.*, *eLife* **4**, e05378 (2015).

ACKNOWLEDGMENTS

P.D., N.D.W., and J.A.K.M. are funded by USAID EPT PREDICT. We acknowledge S. J. Anthony, C. J. Chrisman, Y. Feferholtz, T. Goldstein, C. K. Johnson, D. Nabarro, K. J. Olival, N. Ross, E. Rubin, R. Waldman, B. Watson, C. Zambrana-Torrelio, attendees of the Rockefeller Foundation–funded Bellagio Center Global Virome Project Workshop August 2016 (www.globalviromeproject.org/who-we-are), members of the GVP Core Group and Steering Committee, and co-leads of the GVP Working Groups for their help refining the concept and this manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6378/872/suppl/DC1

10.1126/science.aap7463

VIEWPOINT: COVID-19

COVID-19 testing: One size does not fit all

To control the pandemic, testing should be considered a public health tool

By Michael J. Mina[1,2,3,4] and Kristian G. Andersen[5,6]

Tests for detecting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were developed within days of the release of the virus genome (*1*). Multiple countries have been successful at controlling SARS-CoV-2 transmission by investing in large-scale testing capacity (*2*). Most testing has focused on quantitative polymerase chain reaction (qPCR) assays, which are capable of detecting minute amounts of viral RNA. Although powerful, these molecular tools cannot be scaled to meet demands for more extensive public health testing. To combat COVID-19, the "one-size-fits-all" approach that has dominated and confused decision-making with regard to testing and the evaluation of tests is unsuitable: Diagnostics, screening, and surveillance serve different purposes, demand distinct strategies, and require separate approval mechanisms. By supporting the innovation, approval, manufacturing, and distribution of simpler and cheaper screening and surveillance tools, it will be possible to more effectively limit the spread of COVID-19 and respond to future pandemics.

Many types of tests are available for COVID-19 for clinical and public health use (see the figure). Testing can be performed in a central laboratory, at the point of care (POC), or in the community at the workplace, school, or home. COVID-19 testing begins with specimen collection. For medical use, a nasopharyngeal swab collected by a health care professional has been used for detection of virus infections. Demands on testing throughput for COVID-19, however, have driven new collection approaches, including saliva and less invasive nasal swabs. COVID-19 tests include molecular tests such as qPCR, isothermal amplification, and CRISPR,

as well as antigen tests that detect SARS-CoV-2 proteins directly. Although rapid antigen tests have lower analytical sensitivity (i.e., require greater amounts of virus material to turn positive) than qPCR-based tests, their ability to detect infectious individuals with culturable virus is as high as for qPCR (*3*). Specificity (i.e., correctly identifying those not infected with SARS-CoV-2) of antigen tests achieves comparable results to molecular tests (*4*).

Diagnostic testing for COVID-19 focuses on accurately identifying patients who are infected with SARS-CoV-2 to establish the presence or absence of disease and is performed on symptomatic patients or asymptomatic individuals who are at high risk of infection. This type of testing requires assays that are highly sensitive, so as to not miss COVID-19 patients (false negatives), and specific, so as to not wrongly diagnose SARS-CoV-2–negative individuals as having COVID-19 (false positives). These tests are typically performed by centralized high-complexity laboratories with specialized equipment using qPCR assays, with results that can be reported within 12 to 48 hours. Major bottlenecks in testing, however, have led to turnaround times exceeding 5 to 10 days in some regions, making such tests useless to prevent transmission.

POC diagnostic testing at medical facilities can be qPCR assays, isothermal amplification, or antigen-based (*4*). These POC tests often require instruments that run a limited number of tests and can return results in under an hour. The need for an instrument limits the number of tests that can be performed and where they can be used. However, newer antigen tests are becoming available that do not require instruments or skilled operators, potentially allowing for much more distributed POC testing.

Surveillance testing of populations can be used both as a tool for understanding historical exposures and as a measure of ongoing community transmission. For the former, serological testing of individuals for the presence of SARS-CoV-2–specific antibodies is used to identify those previously infected. For the latter, surveillance testing can be an effective way to monitor real-time SARS-CoV-2 spread in communities. One promising method is wastewater surveillance, which has been used to assess community transmission of polio virus (*5*) and has shown potential for COVID-19 (*6*). qPCR testing of wastewater is used to detect SARS-CoV-2, and frequency dynamics of viral genetic material indicate COVID-19 infections in a community. Surveillance can also be performed from swab or saliva samples taken directly from individuals, and, in populations with low COVID-19 prevalence, pooling can be used to increase capacity and lower cost.

For surveillance testing, the goal is not identification of every case but rather the collection of data from representative samples that accurately measure prevalence and serve to inform public health policy and resource allocation. Because the focus is on extrapolations to the population and not the individual, tests with known deviations from 100% sensitivity and specificity are still appropriate when the variance can be statistically corrected (*7*). To be most effective, results should include reported qPCR cycle thresholds, which is an estimate of viral load (*7*), to model epidemic trajectory and allow for real-time evaluation of mitigation programs (*8*), including once vaccination programs have begun.

Screening testing of asymptomatic individuals to detect people who are likely infectious has been critically underused yet is one of the most promising tools to combat the COVID-19 pandemic (*9*). Infection with SARS-CoV-2 does not lead to symptoms in ~20 to 40% of cases, and symptomatic disease is preceded by a presymptomatic incubation period (*10*). However, asymptomatic and presymptomatic cases are key contributors to virus spread, complicating our ability to break transmission chains (*10*).

Entry screening to detect infectious individuals before accessing facilities (e.g., nursing homes, restaurants, and airports), along with symptom screening and temperature checks, can be beneficial, particularly in high-risk facilities such as skilled nursing facilities. When used strategically, entry-screening measures can be effective at suppressing transmission. Entry screening requires testing that provides rapid results—ideally within 15 min—to be most effective. The required sensitivity and specificity of entry-screening tests are, like all tests, context dependent. Entry-screening tests for a nursing home, for example, must be highly sensitive because the consequences of bringing SARS-CoV-2 into a nursing home can be devastating. Such tests must also be highly specific because the consequences of grouping a false-positive person with COVID-19–positive individuals could be deadly. Conversely, because children have substantially reduced mortality from COVID-19, entry screening into schools might require greater compromise that balances resources and sensitivity to test as many individuals as possible with a need to minimize disruptive false positives. Key to use of tests for entrance screening is that a negative test alone should not be considered sufficient to enter—that should be based on satisfying other requirements, including masks and physical distancing. Conversely, a positive test should be sufficient to bar entry in most settings.

Public health screening is potentially the most powerful form of COVID-19 testing, aimed at outbreak suppression through maximizing detection of infectious individuals. This type of screening entails frequent serial testing of large fractions of the population, through self-administered at-home rapid tests, or in the community at high-contact settings, such as schools and workplaces (*9*). Public health screening can achieve herd effects by stopping onward spread through detection of asymptomatic or presymptomatic cases (fig. S1).

Notably, not every transmission chain needs to be severed to achieve herd effects. Mathematical models that incorporate relevant variation in viral loads and test accuracy suggest that with frequent testing of a large fraction of a population, a sufficient number of cases could be detected to create herd effects (*11*). For example, Slovakia undertook public health screening to address COVID-19 (*12*): During a 2-week period, ~80% of the population was screened using rapid antigen tests. With 50,000 cases identified, combined with other public health measures, it reduced incidence by 82% within 2 weeks (*12*). An important feature of large-scale public health screening is that centrally controlled reporting and contact tracing programs are not essential to induce herd effects as they are for surveillance testing. In a robust public health screening program, sufficient numbers of people are routinely testing themselves, such that contact tracing is subsumed by the screening program (*11*).

Similar to home pregnancy tests, screening tests should be easy to obtain and administer, fast, and cheap. Like diagnostic tests, these tests must produce very low false-positive rates. If a screening test does not achieve high-enough specificity (e.g., >99.9%), screening programs can be paired with secondary confirmatory testing. Unlike diagnostic tests, however, the sensitivity of screening tests should not be determined based on their ability to diagnose patients but rather by their ability to accurately identify people who are most at risk of transmitting SARS-CoV-2. Such individuals tend to have higher viral loads (*13*), which makes the virus easier to detect (*14*). A focus on identifying infectious people means that frequency and abundance of tests should be prioritized above achieving high analytical sensitivity (*11*). Indeed, loss in sensitivity of individual tests, within reason, can be compensated for by frequency of testing and wider dissemination of tests (*9*). In addition, public health messaging should en-

[1]Center for Communicable Disease Dynamics, Departments of Epidemiology and Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Department of Pathology, Clinical Microbiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. [6]Scripps Research Translational Institute, La Jolla, CA, USA.
Email: mmina@hsph.harvard.edu; andersen@scripps.edu

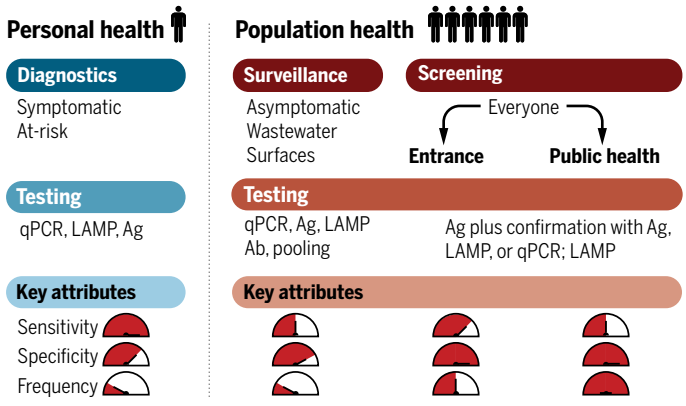
sure appropriate expectations of screening, particularly around sensitivity and specificity so that false negatives and false positives do not erode public trust.

Tests for public health screening require rapid, decentralized solutions that can be scaled for frequent screening of large numbers of asymptomatic individuals. Lateral-flow antigen tests and upcoming paper-based synthetic biology and CRISPR-based assays fit these needs and could be scaled to tens of millions of daily tests (*9*). These tests are simple and cheap, can be self-administered, and do not require machines to run and return results. The Abbott BinaxNOW rapid antigen test, which recently received an Emergency Use Authorization (EUA) in the United States as a diagnostic device, also comes with a smartphone app, allowing self-reporting of COVID-19 status that could be used instead of centralized reporting by public health agencies. Critically, despite being shown to be highly effective at detecting infectious individuals (*14*), very few of these tests are currently approved for screening of asymptomatic individuals, substantially limiting their utility. If such tests were made available direct to consumer (priced to allow equitable access) or produced and provided free of charge by governments, individuals could obtain their COVID-19 status at their own choosing and without complex medical decisions.

Testing is a central pillar of clinical and public health response to global health emergencies, including the COVID-19 pandemic. Nearly all testing modalities have a role, and the one-size-fits-all approach to testing by many Western countries has failed. Many lower- and middle-income countries—including Senegal, Vietnam, and Ghana—have fared far better in their COVID-19 response, often using strong testing programs. The focus on diagnostic tests and the use of preexisting authorization pathways focused on qPCR-based clinical diagnostics not only slows the development and deployment of new surveillance and screening tests but also confuses the picture of what metrics effective public health tools should achieve. Testing to diagnose a patient with COVID-19 is fundamentally different from testing a person to prevent onward transmission. Regulatory pathways should be modified to incorporate these differences so that public health and screening tests are appropriately evaluated. It is necessary to be innovative and produce, distribute, and continuously improve the tests that exist to save lives and gain control of the COVID-19 pandemic.

COVID-19 testing strategies

Testing for SARS-CoV-2 can be for personal or population health. Collection can be from symptomatic or asymptomatic individuals, as well as from wastewater and swabs of surfaces. The tests may be performed in central laboratories, at the POC, or using rapid tests. Attributes of tests differ according to application.



Ab, antibody; Ag, antigen; LAMP, loop-mediated isothermal amplification; POC, point of care; qPCR, quantitative polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. V. M. Corman *et al.*, *Euro. Surveill.* **25**, 2000045 (2020).
2. M. G. Baker *et al.*, *N. Engl. J. Med.* **383**, e56 (2020).
3. A. Pekosz *et al.*, *medRxiv* 10.1101/2020.10.02.20205708 (2020).
4. R. Weissleder *et al.*, *Sci. Transl. Med.* **12**, abc1931 (2020).
5. H. Asghar *et al.*, *J. Infect. Dis.* **210**, S294 (2014).
6. A. Nemudryi *et al.*, *Cell Rep. Med.* **1**, 100098 (2020).
7. R. Kahn *et al.*, *medRxiv* 10.1101/2020.05.02.20088765 (2020).
8. J. A. Hay *et al.*, *medRxiv* 10.1101/2020.10.08.20204222 (2020).
9. M. J. Mina *et al.*, *N. Engl. J. Med.* **383**, e120 (2020).
10. X. He *et al.*, *Nat. Med.* **26**, 672 (2020).
11. D. B. Larremore *et al.*, *Sci. Adv.* 10.1126/sciadv.abd5393 (2020).
12. M. Pavelka *et al.*, "The effectiveness of population-wide, rapid antigen test based screening in reducing SARS-CoV-2 infection prevalence in Slovakia," CMMID Repository, 11 November 2020; https://bit.ly/36BkxV5.
13. E. A. Meyerowitz *et al.*, *Ann. Intern. Med.* 10.7326/M20-5008 (2020).
14. V. M. Corman *et al.*, *medRxiv* 10.1101/2020.11.12.20230292 (2020).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6525/126/suppl/DC1

Published online 21 December 2020
10.1126/science.abe9187

GLOBAL HEALTH

EDITORIAL

Antimicrobial resistance: A global challenge

Simon J. Howard is the Public Health Specialty Registrar in the Office of the Chief Medical Officer, Department of Health, London, UK, and Associate Member of the Fuse Centre for Translational Research in Public Health at Newcastle University. E-mail: sjhoward@doctors.org.uk

IN 1945, ALEXANDER FLEMING SHARED THE NOBEL PRIZE FOR PHYSIOLOGY OR MEDICINE in recognition of his part in the discovery of penicillin. In his Nobel Lecture, he warned that "it is not difficult to make microbes resistant to penicillin," and that people may die as a result of bacteria becoming resistant to antibiotics in the future. Fleming's prediction has come true. Estimates suggest that around 25,000 people in Europe (http://bit.ly/1hi4oR7) and 23,000 people in the United States (https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf) die each year from infections caused by bacteria that are resistant to antibiotics. Meanwhile, in Asia infections caused by resistant bacteria kill one child every five minutes. The problem of antimicrobial resistance can only be solved through concerted global action. Antimicrobials include antibiotics to combat bacterial infections as well as treatments for infections caused by viruses and fungi. The greatest threat is the rapid growth of resistance to our current armamentarium of antibiotics.

The development of antimicrobial resistance is due to the natural Darwinian process of evolution. When bacterial cells are exposed to antibiotics, those cells without natural resistance to the drugs will

Sarah Hopwood is the Assistant Private Secretary in the Office of the Chief Medical Officer, Department of Health, London, UK.

Sally C. Davies is the Chief Medical Officer for England.

Citation:
S. J. Howard, S. Hopwood, S. C. Davies, Antimicrobial resistance: A global challenge. *Sci. Transl. Med.* **6**, 236ed10 (2014).

10.1126/scitranslmed.3009315

be killed; those with resistance will thrive in an environment of reduced competition. Given repeated antibiotic exposure and time, any bacterial species will develop some form of resistance. Society's misuse of antibiotics over the last century has exacerbated and expedited this process, and turned resistance into a current and grave threat to modern medicine. Without antibiotics, the infection-related risks of procedures—from surgery to organ transplants—will substantially increase mortality rates. Treating cancer patients on chemotherapy with antibiotics to prevent infection will become impossible.

The economic cost of allowing antimicrobial resistance to continue is likely to be extreme. The costs associated with antimicrobial resistance are already large, and will grow exponentially if the problem is not tackled. Costs will spiral as we sustain longer hospital admissions and resort to the few newer, more expensive antibiotics. Annual health care costs and productivity losses attributable to antimicrobial resistance are difficult to calculate and estimates vary considerably, yet all the estimates are large. One estimate suggests that the total societal and health care cost in Europe is €1.5 billion, and another estimate suggests that the total cost in the United States is $55 billion (http://bit.ly/1hi4oR7; https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf).

Immediate and coordinated global action is required to preserve the effectiveness of the current antimicrobial armoury (often referred to as "stewardship") and to incentivise the development of new antimicrobials. Preservation of antibiotic effectiveness begins at the individual level. Throughout society, assiduous hygiene practices need to be reinforced. The effect of regular handwashing in health care settings on incidence rates of infections caused by resistant bacteria is well-documented (http://bit.ly/1hi7PXS). In the community, improved sanitation and hygiene practices are likely to reduce both the burden of infection (and hence the use of antibiotics) as well as the spread of resistant organisms.

Individual action is also required on the part of prescribers. A complex web of influence drives prescribing behavior, including demands from patients, competition from alternative systems of health care, and in some countries, financial incentives to prescribe. To reduce the likelihood of the development of resistance, we must ensure that prescribers offer the right antibiotic, at the right dose, for the right amount of time. This includes narrowing the spectrum of antibiotic therapy in a timely manner as microbiological evidence of the causative agent becomes available. This could be aided through the development of rapid and cheap point-of-care diagnostics, which may help to reduce the incidence of initial prescriptions for broad-spectrum antibiotics.

Governments around the world need to tackle the gross misuse of antibiotics. For example, most of the 100,000 to 200,000 metric tons of antibiotics manufactured every year is used in the agricultural, aquacultural, and veterinary sectors (http://bit.ly/1hi9igO). A pricing paradox exists in farming in which antibiotics, a scarce natural resource, cost less than implementation of more effective hygiene practices. Indeed, antibiotics are used in many countries to promote animal growth and hence increase meat yield, rather than for the direct treatment of infections. Reduction of antibiotic use in animal rearing must be achieved while maintaining the security of the food supply. We need to better understand the link between animal antibiotic use and human health, and the costs of inaction to society, the economy, and our health. International collaboration on the introduction, surveillance, and enforcement of guidelines regarding the appropriate use of antibiotics could be fostered through a resolution passed by the Food and Agriculture Organisation of the United Nations (www.fao.org/home/en).

In many countries, an over-the-counter supply of antibiotics without a prescription is the norm. This greatly reduces control of the antibiotic supply, as well as our ability to monitor the use of antibiotics and the development of resistance. However, in countries where access to the advice of qualified medical practitioners is limited, such supplies probably save many lives. There is a balance to be struck between limiting the availability of antibiotics and ensuring timely access to treatment for those who need them. Such a balance is difficult to strike in countries with robust health care systems, but is exponentially more difficult in those with limited health facilities. The increasing pattern of resistance shown by the Enterobacteriaceae, the leading cause of newborn sepsis in developing countries, demonstrates that resistance is a serious threat in these countries. Concerted global action is required to tackle this problem and also to ensure equity of access to effective treatments. Above all, it is important that the needs of all countries, regardless of income or geographical location, are considered as plans to tackle antimicrobial resistance are developed.

To this end, the World Health Organization (WHO) executive board adopted a resolution on antimicrobial resistance in January of this year. If passed by member states at the World Health Assembly next week, the resolution will ask the WHO to develop a global action plan to tackle this major public health issue. In addition, passage of the resolution will mean that the WHO will support individual countries in developing their own strategies for dealing with antimicrobial resistance. This resolution is an important step toward tackling antimicrobial resistance worldwide.

There is also a pressing need to address the drug-development pipeline. We have previously been fortunate enough to have a continued supply of antibiotics to replace those to which bacteria have become resistant, but there has not been a new class of antibiotic developed since 1987 (http://bit.ly/1hibxAM).

The development of new antibiotics requires innovative funding solutions to promote research in this field while removing the incentive for pharmaceutical companies to attempt to maximize sales volume of any newly developed product. Pharmaceutical companies are currently not incentivized to focus on antibiotics compared with other drugs for long-term conditions, such as statins and antihypertensives for cardiovascular disease, because antibiotics are given for shorter periods of time and therefore are far less profitable. Drugs for chronic conditions are taken every day for the rest of a patient's life, whereas antibiotics may only be taken for a week once every few years. New models need to be developed such as decoupling sales volume from monetary reward. Such models may be public-private partnerships, advanced purchase agreements, or something entirely new and are likely to require collaborations among academia, research funders, and not-for-profit organizations. Because of the international nature of regulations and in order to secure widespread availability of new products and foster global cooperation, any economic model developed would need to be standardized across nations. As such, the model will need to be acceptable to all nations.

Antimicrobial resistance is of such concern to the UK that it is now on the risk registers for both the Department of Health and the Department for Environment, Food and Rural Affairs, alongside threats such as climate change. These risk registers assess the likelihood and potential impact of major risks that may directly affect the UK within the next 5 years and allow the relevant departments to plan their response to such events. Antimicrobial resistance, however, can only be successfully addressed through international cooperation. Given that resistant bacteria can be easily transmitted between countries, any mechanism that incentivizes the development of new antibiotics will only be effective if applied to the international pharmaceutical market as a whole. New basic and translational research efforts will be needed to identify new microbial targets and to develop drugs against them. We also need to galvanize international collaborations to prevent the global spread of antimicrobial resistance. This includes prevention and control, surveillance and monitoring, as well as stewardship and conservation. By working together as countries, we believe that sustainable solutions can be developed. Passage of WHO's resolution on antimicrobial resistance at the World Health Assembly next week would be an important first step in helping to solve this international crisis.

– Simon J. Howard, Sarah Hopwood, Sally C. Davies

Acknowledgments: Sally C. Davies is Chair of the WHO Science and Technical Advisory Group on Antimicrobial Resistance and the Antimicrobial Resistance Forum of the World Innovation Summit for Health.

EPIDEMIOLOGY

Modeling infectious disease dynamics in the complex landscape of global health

Hans Heesterbeek,[1]*† Roy M. Anderson,[2] Viggo Andreasen,[3] Shweta Bansal,[4] Daniela De Angelis,[5] Chris Dye,[6] Ken T. D. Eames,[7] W. John Edmunds,[7] Simon D. W. Frost,[8] Sebastian Funk,[4] T. Deirdre Hollingsworth,[9,10] Thomas House,[11] Valerie Isham,[12] Petra Klepac,[8] Justin Lessler,[13] James O. Lloyd-Smith,[14] C. Jessica E. Metcalf,[15] Denis Mollison,[16] Lorenzo Pellis,[11] Juliet R. C. Pulliam,[17,18] Mick G. Roberts,[19] Cecile Viboud,[18] Isaac Newton Institute IDD Collaboration‡§

Despite some notable successes in the control of infectious diseases, transmissible pathogens still pose an enormous threat to human and animal health. The ecological and evolutionary dynamics of infections play out on a wide range of interconnected temporal, organizational, and spatial scales, which span hours to months, cells to ecosystems, and local to global spread. Moreover, some pathogens are directly transmitted between individuals of a single species, whereas others circulate among multiple hosts, need arthropod vectors, or can survive in environmental reservoirs. Many factors, including increasing antimicrobial resistance, increased human connectivity and changeable human behavior, elevate prevention and control from matters of national policy to international challenge. In the face of this complexity, mathematical models offer valuable tools for synthesizing information to understand epidemiological patterns, and for developing quantitative evidence for decision-making in global health.

Thirty-five years ago, it was believed that the health burden of infectious diseases was close to becoming insignificant as hygiene, improved nutrition, drugs, and vaccines brought about a steady decline in overall mortality (*1*). In recent decades, however, it has become clear that the threat from serious infectious diseases will persist, and human mortality attributed to infection is projected to remain at current levels of 13 to 15 million deaths annually until at least 2030 (*2*). Successes in eradicating smallpox and rinderpest have been isolated events in a landscape of endemic and epidemic infections (*3*). Newly emerging infectious agents represent a continuing challenge—for example, HIV in the 20th century; more recently, severe acute respiratory syndrome (SARS) and Middle Eastern respiratory syndrome (MERS) coronaviruses; West Nile Virus; Nipah virus; drug-resistant pathogens; novel influenza A strains; and a major Ebola virus outbreak in 2014–2015. Most new infections enter the human population from wildlife or livestock, and the possibilities

Originally published 13 March 2015

for emergence and spread in the coming decades are likely to increase as a result of population growth, increased urbanization and land changes, greater travel, and increased livestock production to meet demands from the world's expanding population (*4*–*8*). In our modern world of instant communication, the changing behavior of individuals in response to publicity about epidemics can have profound effects on the course of an outbreak (*9*, *10*). Phylogenetic data shed light on an additional layer of complexity (*11*), as will increased understanding of the human genome in relation to susceptibility, infectiousness, and its duration. At the same time, the development of effective new vaccines remains a difficult challenge, especially for antigenically very variable pathogens (e.g., HIV or falciparum malaria) and for pathogens that stimulate immunity that is only partly protective (e.g., *Mycobacterium tuberculosis*) or temporary (e.g., *Vibrio cholerae*).

In the face of this complexity, computational tools (Box 1) are essential for synthesizing information to understand epidemiological patterns and for developing and weighing the evidence base for decision-making. Here, we review the contribution of these tools to our understanding of infectious disease dynamics for public health by using representative examples and by ranging into current developments. We argue that to improve decision-making for human health and for sustaining the health of our food systems, experts on infectious disease dynamics and experts on prevention and control need to collaborate on a global scale. To succeed, quantitative analysis needs to lie at the heart of public health policy formulation.

Models and public health policy formulation

The value of mathematical models to investigate public health policy questions was recognized at least 250 years ago when, in 1766, Daniel Bernoulli published a mathematical analysis of the benefits of smallpox inoculation (then called variolation) (*12*). In the past 50 years, the study of infectious disease dynamics has grown into a rich interdisciplinary field. For example, decision-making for vaccination strategies increasingly depends on model analyses in which infection dynamics are combined with cost data (Box 2, Influenza: prevention and control). In recent decades, responses to major infectious disease outbreaks, including HIV, bovine spongiform encephalopathy (BSE), foot-and-mouth disease (FMD), SARS, and pandemic and avian influenza, have shown both the need for and capabilities of models (Box 3, HIV: Test and treat strategy). Model-based analysis of such outbreaks also continually brings improvements in methodology and data, emerging from the comparison of model prediction with observed patterns.

For infectious agents important to public health, a series of principles has emerged for modeling infection dynamics (Table 1 and Box 4). The basic reproduction number R_0, for example, is a central concept characterizing the average number of secondary cases generated by one primary case in a susceptible population. This concept highlights what must be measured to interpret observed disease patterns and to quantify the impact of selected control strategies (Fig. 1).

Two fundamental properties of the world that shape infectious disease dynamics make computational tools key for understanding reality. The world is essentially a stochastic and highly nonlinear system. The nonlinearity derives not only from the complex interaction between factors involved in transmission, but also from the influence that the infection process has on the distribution of important characteristics at various temporal and spatial scales. This effect is seen in the age-related nature of infection and mortality in HIV changing the age distribution of the population, and in previous exposure to strains of influenza altering the distribution of influenza susceptibility. Such feedback mechanisms contribute to the nonlinearity of infection processes. Nonlinearity also leads to counter-intuitive phenomena (Fig. 2) and prevents simple extrapolation of experience from one situation to another, such as when deciding whether to implement a vaccination policy in different countries (Fig. 1). Mathematical tools, relating to data and processes on a large range of interacting scales, have become essential to explore, anticipate, understand, and predict the effects of feedbacks within such complex systems, including changes caused by intervention.

Current and future opportunities for models in public health

Over the past decade, key public health questions, ranging from emergence to elimination, have posed a range of challenges for modeling infectious disease dynamics, many of which rely on leveraging disparate data sources and integrating data from a range of scales from genomics to global circulation. Given commonalities in processes across pathogens, progress made in one area can lead to advances in another. Progress in the areas described above all build on and inform each other, making this a dynamic time for research in the discipline (*13*). A few themes are chosen to illustrate current trends in model development and public health application.

Real-time outbreak modeling: The Ebola 2014–2015 outbreak

The 2014–2015 outbreak of Ebola in West Africa serves to highlight both opportunities and challenges in modeling for public health. In the initial phase of this outbreak, real-time estimates of the reproduction number or simple exponential extrapolation (*14*) allowed short-term predictions of epidemic growth that were used, for example, to plan for necessary bed capacity. Quantitative phylogenetic tools applied to samples from initial victims provided important estimates of the origin of the outbreak (*15*). Early mechanistic models that explicitly took into account the roles played by different transmission routes or settings were informed by analysis of earlier outbreaks (*16*, *17*). When the failure to contain the epidemic with methods successful in previous outbreaks led to a scale-up of capacity driven by international aid, such models were used to assess the impact of, for example, reducing transmission at funerals (*17*) and whether the construction of novel types of treatment centers could end up doing more harm than good. Ensuring that the most effective combinations of interventions were implemented required close and fast interaction between modelers and policymakers (*18*). Looking forward, models are now used to help clinical trial design and inform a debate on the optimal deployment of initially

[1]Faculty of Veterinary Medicine, University of Utrecht, Utrecht, Netherlands. [2]School of Public Health, Imperial College, London, UK. [3]Roskilde University, Roskilde, Denmark. [4]Georgetown University, Washington, DC, USA. [5]MRC Biostatistics Unit, Cambridge, UK. [6]WHO, Geneva, Switzerland. [7]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene Tropical Medicine, London, UK. [8]University of Cambridge, Cambridge, UK. [9]School of Life Sciences, University of Warwick, UK. [10]School of Tropical Medicine, University of Liverpool, UK. [11]Warwick Mathematics Institute, University of Warwick, Coventry, UK. [12]Department of Statistical Science, University College London, London, UK. [13]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [14]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. [15]Department of Zoology, University of Oxford, Oxford, UK, and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [16]Heriot-Watt University, Edinburgh, UK. [17]Department of Biology–Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA. [18]Division of International Epidemiology and Population Studies, Fogarty International Center, NIH, Bethesda, MD, USA. [19]Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand.

***Corresponding author. E-mail: j.a.p.heesterbeek@uu.nl**

†Authors, apart from first and last author, are in alphabetical order. ‡All authors are members of this collaboration. §In addition to the authors listed above, this collaboration includes: Nimalan Arinaminpathy,[1] Frank Ball,[2] Tiffany Bogich,[3] Julia Gog,[4] Bryan Grenfell,[3] Alun L. Lloyd,[5] Angela Mclean,[6] Philip O'Neill,[2] Carl Pearson,[11] Steven Riley,[1] Gianpaolo Scalia Tomba,[12] Pieter Trapman,[13] James Wood[7]. Affiliations: [1]Imperial College, London, UK. [2]University of Nottingham, Nottingham, UK. [3]Princeton University, Princeton, NJ, USA. [4]University of Cambridge, Cambridge, UK. [5]North Carolina State University, Raleigh, NC, USA. [6]University of Oxford, Oxford, UK. [11]University of Florida, Gainesville, FL, USA. [12]University of Rome "La Sapienza," Rome, Italy. [13]University of Stockholm, Stockholm, Sweden.

scarce Ebola vaccines, once such vaccines become available.

With the opportunities of real-time modeling for public health come specific challenges. The imperative to produce reliable and meaningful analysis for those treating infected people has to be balanced against the pressures and delays of scientific publication. In an ongoing outbreak, data can be patchy and reporting delayed, and different data sources are not always synthesized. When the Ebola outbreak expanded explosively in the summer of 2014, data were often lacking on the effect on transmission dynamics of the various control measures that operated simultaneously in the hectic circumstances of the most severely hit areas. In any emerging epidemic, underreporting is a critical challenge for ongoing assessment of this epidemic and has had enormous impact on predictions of out-break size, but also of outbreak impact—for example, in terms of the case-fatality ratio (the proportion of cases that lead to death). Early in any outbreak, this estimate of severity can suffer from imprecise information on both the numerator (if not all deaths due to the infection are identified as such; for example, because health services are overwhelmed caring for the sick) and the denominator (if cases are not reported or, conversely, noncases get reported as cases if they are not laboratory-confirmed). This caused problems early in the H1N1 influenza outbreak first reported in in Mexico in 2009, as well as in the current Ebola outbreak. Although level of underreporting can be estimated from retrospective serological studies, it is usually not identifiable in real-time data.

These limitations make it almost impossible to make reliable long-term predictions. Thus, modeling results are often based on scenarios in which a pathogen spreads unaltered by behavioral changes or the public health response. This rarely reflects reality, especially in such a devastating outbreak as Ebola, where the situation constantly changes owing to growing awareness in the community, as well as national and international intervention. Careful communication of findings is key, and data and methods of analysis (including code) must be made freely available to the wider research community. Only in this way can reproducibility of analyses and an open exchange of methods and results be ensured for maximal transparency and benefit to public health.

Emergence of novel human pathogens

There is an ever-present hazard that novel human pathogens emerge from livestock and wild mammal and bird reservoirs. Research on potential emerging zoonoses draws on concepts from across the spectrum of infectious disease dynamics, disease ecology, microbiology, and phylogenetic analysis. Particular challenges include estimating human-to-human transmissibility against a backdrop of ongoing zoonotic spillover, detecting anomalous outbreaks, and assessing the risk that more dangerous strains may arise through pathogen evolution.

The recently identified gap in methodology for zoonoses with weak human-to-human transmission (*6*) is being filled with new approaches for estimating R_0 and other transmission-related quantities from subcritical outbreak data (*19*–*21*). These studies address key public health concerns, but rely on strong assumptions regarding the quality and completeness of case observations. Better information on surveillance program efficacy could be gained through serological surveys (where blood and saliva samples reveal evidence of past and present infections) or sociological study, and modeling studies can help to design and characterize efficient surveillance programs (*22*). Given the predominance of zoonotic pathogens among emerging infections, models for transmission dynamics and evolution in multispecies ecosystems and food webs (consisting of host species and nonhost species interacting ecologically and epidemiologically) are a crucial area for future development (*6*, *23*). The greatest challenge—and the greatest prize—in modeling emerging zoonoses is to assess which diseases pose the most risk to humans and how these might change over time and in different localities (*24*). Such tasks, which will join molecular studies to experimental infections to epidemiological and ecological surveys, will drive empirical and theoretical efforts for decades to come.

The rising availability of pathogen genome sequence data, coupled with new computational methods, presents opportunities to identify with precision "who infects whom" and the networks of infection between humans and reservoirs (*25*). Full realization of this potential, though, will require denser and more systematic whole-genome sampling of pathogens coupled with associated epidemiological data, as well as baseline information on genetic diversity and evolutionary rates, especially in animal hosts (*26*).

Pathogen evolution and phylodynamics

As pathogen genetic data become increasingly available, modelers are finding ways to synthesize these new data streams with more traditional epidemiological information in phylodynamic tools (*27*, *28*). However, current frameworks employ compartmental epidemiological models, which

Box 1. Quantitative tools in infectious disease dynamics.

Here, we use the words "computational tools" loosely. In infectious disease dynamics, there is a broad range of relevant quantitative tools, and we refer to the entire collection. It comprises statistical methods for inference directly from data, including methods to analyze sequencing and other genetic data. This leads to estimates of important epidemiological information such as length of latency, incubation and infectious periods and their statistical distributions, inferred transmission chains and trees early in outbreaks, the risks related to various transmission routes, or estimates of rates of evolution. Mathematical models in the strict sense refer to mathematical descriptions of processes thought to be associated with the dynamics of infection—for example, in a population or within an individual. Such models take many forms, depending on the level of biological knowledge of processes involved and data available, and depending on the purpose.

Several classes of model are used, spanning the spectrum of information available. At one end of the range are detailed individual-based simulation models, where large numbers of distinct individual entities (with their own characteristic traits such as age, spatial location, sex, immune status, risk profile, or behavior pattern) are described in interaction with each other, possibly in a contact network, and with the infectious agent. At the other end are compartmental models where no individuals are recognized, but only states for individuals (for example: susceptible, infectious, immune) aggregated into compartments where everyone has the same average characteristics and where interaction is typically uniform (everybody interacts with everybody else). Such models do not describe the disease history of single individuals, but rather the time evolution of aggregated variables, such as the number of individuals that are currently susceptible.

Mathematical models can have both mechanistic parts in their description, based on assumptions about biological mechanisms involved, and more phenomenological parts, where there is a statistical or presumed relation between variables, without clear assumptions from which this relation can be derived. An example of the former is the assumption of mass action to describe interaction between individual hosts; an example of the latter is an empirical relation between the length of an infectious period in a mosquito and environmental temperature.

For infectious disease dynamics, our world is clearly stochastic, in that chance events play a role in many of the processes involved. Certainly at lower levels of biological aggregation, chance dominates—for example, in infection of individual cells or in contacts individual hosts make. At higher aggregation levels, many cells or individuals interact, and chance effects may average out to allow deterministic descriptions. There are purely stochastic models, purely deterministic models, and models that are mixed. It is important to point out that, even though the world is stochastic, stochastic descriptions are not by definition better than deterministic descriptions. Both are still models of reality, and the fact that chance plays a role may have a far less significant influence on model outcome and prediction than choices made in the relations between ingredients and variables.

Areas of rapid growth are statistical and numerical methods and tools to estimate model parameters from, often scarce, mismatched or incomplete data, and to contrast model output with real-world observations.

do not make efficient use of individual-level epidemiological data. Although sampling theory is well developed for standard surveillance data, the relationship between a set of pathogen sequences and the phylogeny inferred from a population sample is more complex (*11*). Many-to-one mapping possibilities between, on the one hand, combinations of epidemiological, immunological, and evolutionary processes shaping sequences and, on the other hand, the inferred phylogeny, demand the integration of diverse data sources and an increased focus on systematic sampling.

Phylodynamic studies to date have largely focused on fast-evolving RNA viruses, driven by the large amount of data generated for clinical [e.g., hepatitis C virus (HCV) or HIV)] or surveillance (e.g., influenza) purposes (*11*). Replicating these efforts on an expanded array of pathogens, including DNA viruses, bacteria, fungi, protozoa (e.g., malaria), and helminths, is a promising avenue for future research (*29*). It is of particular importance in the context of the evolution and spread of drug-resistant variants and vaccine escape mutants. However, genome-wide pathogen data also present challenges, in particular in relation to accommodating recombination, reassortment, and mobile genetic elements. Analysis of bacterial genomes usually considers only those genes that are shared across taxa, but there are good reasons to believe that noncore genes play an important role in bacterial evolution, including the evolution of antibiotic resistance (*30*).

Although sequence data are extremely valuable, to link these data fully to disease dynamics, it will be important to determine how sequence changes affect functions related to pathogen fitness, such as replication rate, transmissibility, and immune recognition. Molecular epidemiological studies often treat pathogen genetic variation as simply reflecting the underlying transmission process, whereas in reality such variation may play an important role in determining transmission dynamics, as exemplified by escape from herd immunity by influenza A virus (*31*).

"Deep" sequencing of pathogens within individual hosts generates information on within-host diversity, resulting from evolution within the host (often in response to drug treatment), or multiple infections. To tackle within-host diversity, models that embed pathogen evolution within a transmission tree are needed. Such models, which cross the within- and between-host scales, are only just becoming analytically and computationally feasible despite being proposed several years ago (*32*). Similarly, although progress has been made in scaling inference from genes to genomes (*33*), scaling inference to large numbers of sequences is lagging far behind.

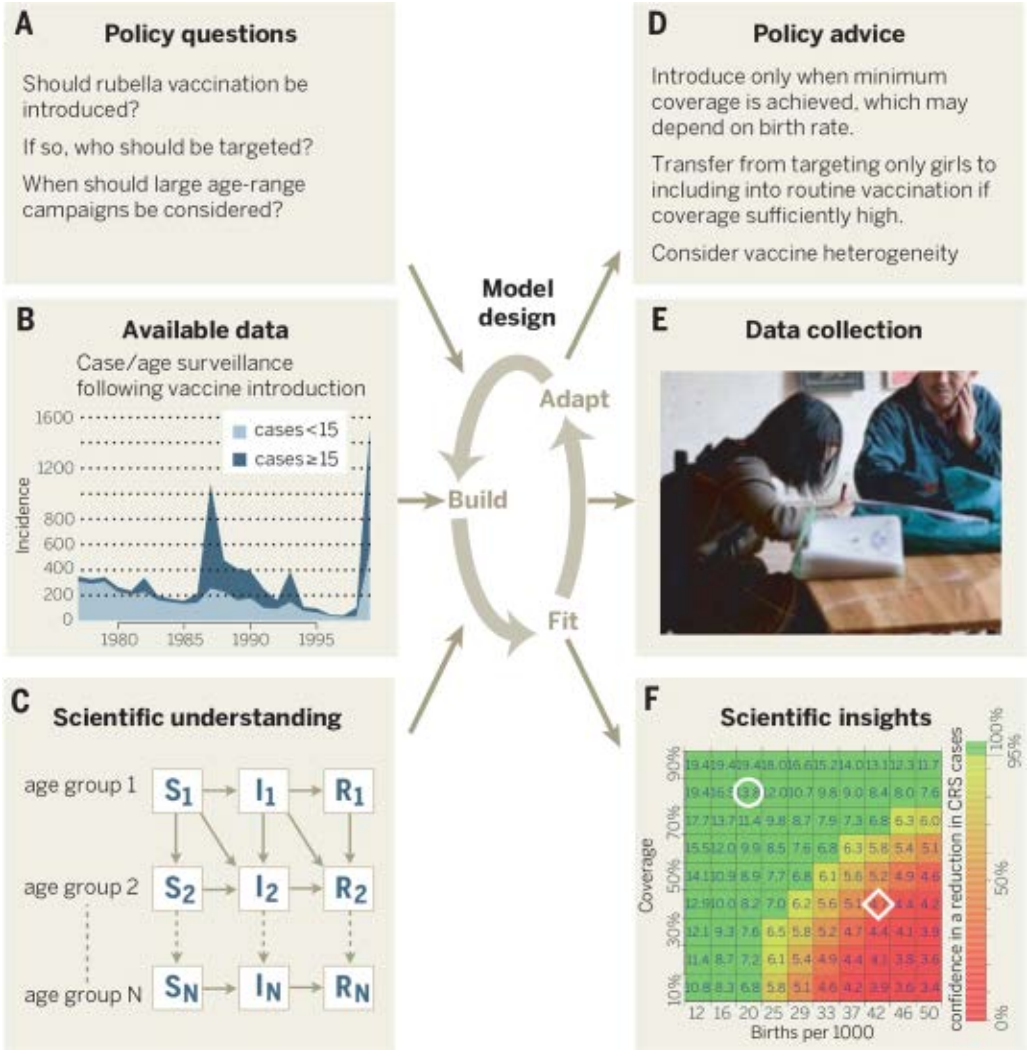


Fig. 1. Modeling for public health exemplified by rubella. (**A** to **F**) Policy questions are formulated; available data are brought to bear on the question. In this example, the incidence of rubella is shown following the introduction of vaccination in individuals under 15 or 15+ years of age in Costa Rica (*127*). Application of a nonlinear age-structured SIR model (see Box 4) to these circumstances led to the collection of key missing data. In the bottom right-hand plot, each square depicts a combination of birth rate and infant vaccine coverage reflecting different countries (e.g., Somalia depicted by a diamond and Nepal by a circle), colored by expected effect on congenital rubella syndrome (CRS) in newborns, related to local R_0 (*128*). This translates into confidence that routine vaccination is likely to reduce the public health burden caused by CRS in Nepal (green), but not in Somalia (red).

Multiple infections

Infectious disease epidemiology evolved by focusing on interactions between a single host species and a single infectious agent. It is becoming increasingly clear that multiple agents simultaneously infecting the same host populations and individuals appreciably add to the public health burden and complicate prevention and control. Coinfections in relation to HIV— for example, tuberculosis and HCV—or coinfection of different strains of influenza A virus raise important public health and evolutionary issues. Multiple agents infecting the same host individual have been shown to influence each other by increasing or decreasing susceptibility and/or infectivity of that individual, thereby influencing the population dynamics of these agents in ways that we have yet to explore and understand (*34*, *35*).

Multiple infections of the same individual with closely related patho-

gens occur when infection elicits no immunity, or only a partial immune response. Macroparasites, including many of the important human helminth infections, are good examples of pathogens that evade human immune responses and cause repeated infection of the same host (*36*). Biological mechanisms giving rise to such multiple infections include sequential reinfections caused by antigenic drift in influenza, antigenic variation in respiratory syncytial virus (RSV), and waning (slow loss of) immunity in pertussis, while lack of cross-protection in many colonizing microparasites— for example, pneumococcus and human papilloma virus (HPV)—allows for multiple concurrent infections. Although the existence of reinfections is a clinical fact, population-level data are scarce as reinfections are often subclinical and individual-based longitudinal infection histories are often only anecdotal. Results from new analytical approaches relating to deep sequencing and neutralization tests covering multiple antigens are being utilized (*37*).

The immunodynamics of influenza have clear policy implications for the identification of high-risk groups in connection with pandemic planning (*38*), while the dynamics of waning immunity are key to the current concerns about

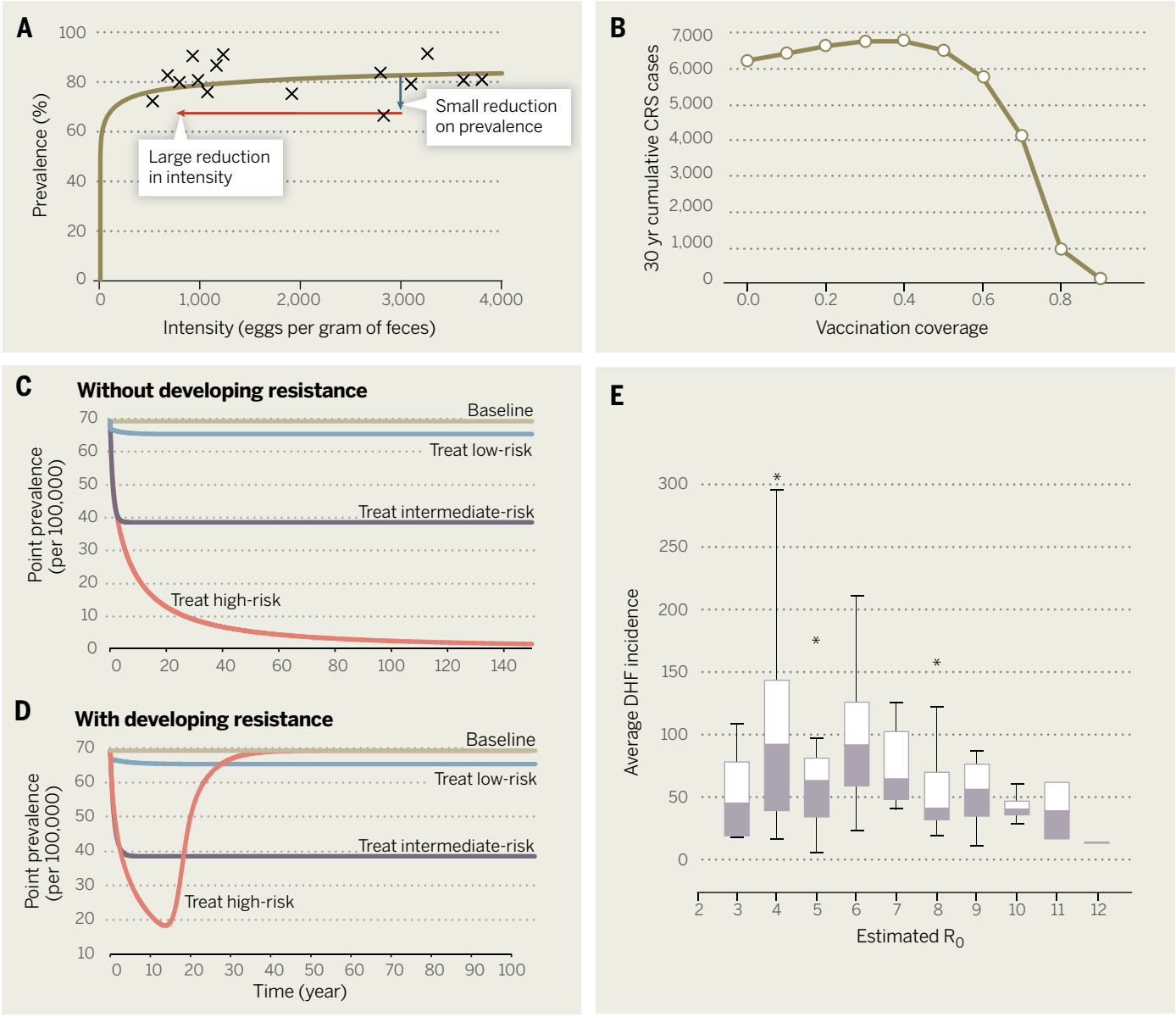


Fig. 2. Examples of counterintuitive effects of nonlinear infection dynamics. (**A**) Nonlinear interaction between prevalence of a helminth infection and infection pressure (as measured by the mean intensity of existing infections) means that control measures must have a disproportionately large impact on intensity before prevalence is reduced. This effect is predicted by a mathematical model (solid line) and corroborated by field data (crosses) (*129*). (**B**) Nonlinear relation between total number of cases of congenital rubella syndrome (CRS) and rubella vaccine coverage, showing that suboptimal levels of vaccine coverage cause worse health outcomes than no vaccination [adapted from (*130*)]. The line shows model predictions; similar effects have been documented for real rubella control situations (*131*). (**C** and **D**) Modeling results of rebound of gonorrhea transmission with different treatment strategies without (C) and with (D) antimicrobial resistance developing [adapted from (*132*)]. In the presence of resistance, focusing treatment on the high-risk core group leads to an increase in prevalence approaching that of untreated baseline prevalence, after an initially strong decline for more than a decade. (**E**) Box plot from field data of a nonlinear relation between R_0 for dengue transmission and average dengue hemorrhagic fever (DHF) incidence across Thailand, showing that control measures that reduce R_0 may paradoxically increase cases of DHF [adapted from (*133*)].

immunization level for pertussis (*39*). Multivalent vaccines covering only a targeted subset from the circulating strains of pneumococcus and HPV pose important new applied problems (*40*). The spread of recombinant viruses implies the existence of multiple infections. One example is the Sydney 2012 strain of norovirus, but how this can occur in such an acute infection remains to be understood, as the time window for multiple exposures is limited, unless subclinical or environmental reservoirs of infection are important. Mathematical models could help to explore how, for example, such subpopulations may contribute to the dynamics of multiple infections.

Behavior of hosts

Human behavior is a fundamental determinant of infectious disease dynamics, whether by affecting how people come in contact with each other, vaccination coverage, reporting biases, or adherence to treatment. Traditional epidemic models have tended to ignore heterogeneity in contact behavior [although early HIV models addressed heterogeneity in sexual behavior by necessity (*41*)]. Increasing sophistication of contact network models (*42*), together with data on epidemiological contacts, creates opportunities for understanding and controlling transmission at a fundamental level (*43*) and opens up the possibility of independent study of relevant social factors (*10*). Recent years have seen exciting developments in the measurement of contact patterns and "who might infect whom" through advances in individual electronic identification technology. This is a promising avenue for linking pathogen genetic data and human behavior.

Contact patterns are not static and can shift during outbreaks as individuals change their behavior in response to perceived risk and public health interventions (*44*). Modeling has illuminated this process—for example, by the incorporation of peer influence on vaccination behavior into models of infectious disease dynamics (*45*, *46*). Analysis of data from online social networks has also created promising opportunities to validate such approaches with empirical observations (*47*, *48*).

Movement and travel are tightly linked to the spread of infection and have been explored through models to highlight commuting and agricultural migration driving local disease transmission (*49*) and global disease patterns through air travel (*50*). These processes are now being investigated to gain insights into the more complex case of vector-borne diseases, such as malaria and dengue, where both host and vector movement can interact to drive local (*51*) and large-scale dynamics (*52*).

Elimination and eradication

Modeling has long provided support for elimination efforts: Vector control (*53*), critical community size (*54*), herd immunity, and critical vaccination threshold (*55*, *56*) were all powerful insights from models framed in relatively simple and homogeneous terms. Subtleties and complexities in many current eradication programs, as well as the availability of novel data sources, have called for a range of extensions in the theory. As we approach elimination targets, disease dynamics have changed in ways that were largely predicted by models, but also in unanticipated ways as a result of ignorance about key epidemiological processes (*3*).

Incentives for control efforts also change, both at the individual level [passive or active refusal to participate can develop (*57*)] and at the country level (*58*). This reinforces the call for development of models of human behavior and its interaction with infectious disease dynamics (*9*) potentially drawing on new data sources from social media (*59*, *60*), as well as for models that can capture national and nongovernmental motivations, interactions, and competition, economical or otherwise. Long-term control puts pathogens under strong selection for resistance, calling for evolution-proof control methods (*61*) and novel vaccine technologies and their optimized delivery (*62*).

Box 2. Influenza: prevention and control.

Human influenza—pandemic and seasonal—remains a major issue in public health owing to the continued emergence of novel genetic strains, and one where models have successfully addressed questions from basic biology to epidemiology and health policy. In recent years, modeling and other quantitative analysis has been used to study at least three major issues: pandemic preparedness and mitigation strategies (*84*–*89*), rethinking vaccination strategies for seasonal influenza (*70*), and improved methods in phylodynamics and influenza strain evolution (*11*). Recent models of influenza fitness have also been developed to predict viral evolution from one year to the next, providing a principled and more precise method for the vaccine selection required every year (*90*).

For seasonal influenza, models have played a key role in providing the scientific evidence base for vaccination policy, making use of the information in multiple, often unavoidably biased, data sources such as syndromic time series, vaccine coverage and efficacy, economic costs, and contact patterns in the population. For example, a combined epidemic and economic model was fitted to fine-grained data from many sources to describe the dynamics of influenza in the United Kingdom, and the influence of previous vaccination programs (*70*). With confidence in the model's predictions based on its ability to capture past patterns, it was used to look at alternative vaccination strategies and led to a new national policy to vaccinate school-age children (*91*). Targeting those individuals most likely to spread the virus, rather than only those most likely to suffer the largest morbidity, is a marked departure from established practice in the UK and is currently under consideration elsewhere (*92*).

Box 3. HIV: test and treat.

Mathematical modeling has played a central role in our understanding of the HIV epidemic, and in informing policy from the outset of our recognition of the pandemic (*93*). Some of the many insights include a model-based analysis of viral load data from inhibition experiments, which revealed the rapid and ongoing turnover of the within-host viral population (*94*), and the use of phylogenetic models to show that the HIV pandemic did not emerge in the 1980s, but had its roots in the early 20th century (*95*).

A key contribution of mathematical modeling has been to identify when viral transmission occurs over the course of infection, which determines the potential to halt spread by various measures. Models have shown that transmission of HIV depends on the epidemic phase and the sexual behavior of the population, and a large proportion of transmissions may occur late in infection (*96*). Model-based inference in the Netherlands also suggested that the effective reproduction number (Box 4) had fallen below 1 due to a combination of low-risk behavior and a very effective diagnosis and treatment program (*97*). The debate was transformed in the mid-2000s, when eradication of HIV through a "test and treat" strategy was hypothesized (*98*, *99*). Subsequent trial results showing that antiretroviral treatment (ART) of HIV-positive individuals could practically eliminate transmission within sexual partnerships when the index case is treated (*100*) have further supported the role of treatment as prevention. Although these findings have not dispelled concerns about transmission early in infection (*93*), or about extra-couple transmission (*101*), it is suggested that high population coverage of ART may have reduced the incidence of HIV infection in rural KwaZulu-Natal, South Africa (*102*).

These findings, combined with the prospect of cheaper, more effective drugs and delivery structures, underpin UNAIDS' goal of "zero new infections" for HIV and the initiation of a multimillion dollar cluster-randomized trial (*103*), which will have its outcome assessed against model predictions. In the meantime, the scientific discussion of the effectiveness of ART in preventing transmission continues, sparked by studies that fail to show a decline in incidence after increased treatment (*104*). Such debates are essential to elucidate areas for improvement of the models used and data needs for such improvement, and to highlight methodological limitations (*105*).

Finally, since the era of smallpox eradication, patterns of global disease circulation have changed radically. Human mobility and migration are increasing global connectivity, strengthening the need for cooperation and international synchronization of efforts (as illustrated by polio). Techniques for analysis of novel data sources are again key here; e.g., mobile phone records provide unique opportunities to understand disease source-sink dynamics (*52*).

Computational statistics, model fitting, and big data

By definition and design, models are not reality. The properties of stochasticity and nonlinearity strongly influence the accuracy of absolute predictions over long time horizons. Even if the mechanisms involved are broadly understood and relevant data are available, predicting the exact future course of an outbreak is impossible owing to changes in conditions in response to the outbreak itself, and because of the many chance effects in play. These stochastic effects dominate developments in situations with rel-

Table 1. Principles for modeling infection dynamics. As different infections have become the focus of public health attention, the modeling community has responded by developing improved concepts and methods. The table concentrates on the period since 1950. The first column lists the classes of infection, and the second column lists factors whose importance to infection dynamics became particularly clear in relation to those infections; the third and fourth columns highlight concepts and methods that were developed in response. For each row, only a few typical references are given. Many factors, concepts, and methods are relevant, in current use, and in continual development for much larger classes of infectious agents.

Motivating studies	Important factors	Concepts	Methods
Malaria [1910s, 1950s onwards, *(53, 106, 107)*]	Transmission via insect vectors; nonlinear dependence of transmission on mosquito biting rate; influence of environmental and climatic variables.	Threshold for control, basic reproduction number.	Models with two host species (host-vector models); using models to support and guide field campaigns; relating models to field data.
Childhood infectious diseases, e.g., measles. [1950s onwards, *(54, 108)*]	Immunizing infections; spatial and temporal heterogeneity; demography; age structure; household structure.	Critical community size and herd immunity; periodic outbreaks; fade out; vaccine efficacy.	*SIR* models; age-structured models; models with periodic forcing; spatial and stochastic models; metapopulation models; time-series models.
Macroparasites. [1970s onwards, *(108)]*	Clumped infections, multistrain and multispecies infections, cross immunity, concurrent infections.	Consequences of overdispersed distribution of parasite load (Fig. 1)	Stochastic models, approximations including hybrid models and moment closure.
Sexually transmitted infections, e.g., HIV. [1980s onwards, *(108, 109)*]	High/low risk groups; nonrandom contact structure; partnerships; within-host strain diversity and evolution; time scale.	Incubation and infectious period distribution; core group; next-generation matrix and operator; partnership dynamics.	Statistical methods (e.g., back calculation); models with (dis)assortative mixing; pair-formation models; within-host dynamic models.
Veterinary outbreaks, e.g. BSE and FMD. [1990s onwards, *(110, 111)*]	Fixed spatial locations with changing contact networks.	Local versus long-range transmission; spatial intervention (ring vaccination/culling); conflict of priorities at different scales.	Individual-based models and spatial simulations (FMD); data-driven real-time modeling; inference of transmission trees.
Novel emerging infections, e.g., SARS, Nipah virus, MERS. [2000 onwards, *(6, 19, 112–115)*]	Behavior change; global interconnectedness and international cooperation in control; responses in absence of biomedical measures; animal reservoirs.	Zoonotic spillover; stuttering chains; importance of index case; superspreaders; unobserved dynamics in an animal reservoir; supershedding	Contact tracing; modeling international spread and control; quarantine and case isolation; individual heterogeneity in infectiousness, incubation, and latency period.
Influenza, including avian influenza. [Present, *(27, 28, 116–118)]*	Distribution of prior immunity; within-population and species strain differences, virus evolution and interaction; role of wildlife and farm animals.	Pandemics; spillover between wild birds and farmed birds; phylodynamics.	Interaction between immunological and epidemiological dynamics; integrating phylogenetic and epidemic methods and models.
Vector-borne diseases, e.g., dengue, malaria. [Present, *(119-122)*]	The influence of climate and environment on vector and pathogen development; animal reservoirs; interaction between strains within-host and between-host.	Dilution effect and role of biodiversity in infectious disease dynamics; reemerging infections.	Evolutionary impact of vaccines/other interventions; synthesis of data from ecology and epidemiology; elimination modeling; statistical modeling of environmental vector suitability.
Bacterial infections, e.g., pneumococcal disease, MRSA, and tuberculosis. [Present, (*39, 123–126*)]	Antibiotic/drug resistance; adaptive dynamics.	Vaccine effectiveness; interacting natural immunity and vaccine boosting.	Modeling interacting and emerging strains; stochastic models in small populations.

atively few infected individuals that occur at emergence, approaching the threshold for sustained host-host spread, or approaching elimination and eradication. This makes it virtually impossible to predict which infectious disease agent is going to emerge and evolve next and where, or to predict when and where the next or last case in an outbreak will occur. There is, typically in complex systems, a fundamental horizon beyond which accurate prediction is impossible. The field has yet to explore where that horizon is and whether computational tools and additional data (and if so which data) can stretch predictions to this limit. In contrast, "what-if" scenarios for public health intervention can provide qualitative (and increasingly semiquantitative) insight into their population consequences.

With growing applications in public health, there is an increasing demand to validate models by making model predictions consistent with observed data. The development of ever-more-powerful computers is accompanied by new techniques utilizing this power, notably for statistically rigorous parameter estimation and model comparison. Techniques such as Markov chain Monte Carlo (MCMC) have become firmly established tools for parameter estimation from data in infectious disease models [e.g., (*63*)], and Monte Carlo based methods will play a pivotal role in addressing the challenges that lie in reconciling predictions and observations (*64*, *65*). Other techniques, such as so-called particle filters, approximate Bayesian computation, emulation, and their combinations with MCMC [e.g., (*66*)], are rapidly developing and allow stochastic models that explicitly account for incomplete observations to be matched to time series of cases, giving insights into scenarios as diverse as cholera in Bangladesh (*67*) and influenza (*68*, *69*). The need to integrate multiple data sources (*70*, *71*), as well as to include uncertainty in model parameters and/or structure, has driven the use of Bayesian approaches.

Although the rapid expansion of infectious disease models and their application over the past decade has coincided with an increase in open access data sets available from a variety of sources, progress in data capture needs to be accelerated. Although some of these technologically advanced data streams have been incorporated into models—for example, to track the incidence of influenza in the United States (*72*), to elucidate the spatial dynamics of measles and malaria in Africa (*53*, *73*), and to chart the spread of dengue globally (*74*)—much more remains to be done to leverage data collected from different sources (e.g., demographic, genetic, epidemiological, treatment, and travel patterns) and at different temporal and spatial scales.

Box 4. Some fundamental terms and concepts.

- **susceptible**: individual who is at risk of becoming infected if exposed to an infectious agent.
- **basic reproduction number, R_0**: average number of infections caused by a typical infected individual in a population consisting only of susceptibles; if $R_0 > 1$, the infectious agent can start to spread.
- **effective reproduction number, R_e**: average number of infections caused by a typical infected individual when only part of the population is susceptible; as long as $R_e > 1$, the agent can continue to spread.
- **herd immunity**: state of the population where the fraction protected is just sufficient to prevent outbreaks ($R_e < 1$).
- **critical elimination threshold, p_c**: proportion of the susceptible population that needs to be successfully protected—for example, by vaccination—to achieve herd immunity; $p_c = 1: 1/R_0$ is a rule of thumb from models when hosts are assumed to mix randomly.
- **force of infection**: per capita rate at which susceptible individuals acquire infection.
- **final size**: fraction of the initial susceptible population that eventually becomes infected during an outbreak.
- **prevalence**: proportion of the population with infection or disease at a given time point.
- **superspreader/supershedder**: infected individual that produces substantially more new cases than the average because of greater infectiousness, longer duration of infectiousness, many more transmission opportunities and contacts, or combinations of these. Even when the average R_0 is relatively small, these individuals have large effects on outbreaks.
- **metapopulation**: collection of populations, separated in space, but connected through movement of individuals.
- **critical community size**: minimum number of individuals in a population that allow an infectious agent to persist without importation of cases.
- **case fatality ratio**: proportion of symptomatic infections that result in death.
- **SIR model**: most basic model metaphor for immunizing infections where each living individual is assumed to be in one of three epidemiological states at any given time: susceptible, infected and infectious, and recovered and immune. The model specifies the rates at which individuals change their state. Individuals progress from S to I when infected, then from I to R upon recovery. Many variants exist—for example, recognizing different classes of S, I, and R individuals, depending on individual traits such as age.

Concluding remarks

Infectious diseases are an important frontier in public health, and their prevention and control call for global, rather than national or regional, coordinated efforts (*75*–*78*). The success of smallpox and rinderpest eradication campaigns shows the possibilities; the global spread of newly emerged pathogens (recently avian influenza strains and MERS coronavirus), the difficulties in curbing the spread of antibiotic resistance, the upsurge of polio toward the "end-phase" of its eradication, and the recent unprecedented spread of Ebola virus, are examples that show the need for international coordination and collaboration. Nonlinearity in infectious disease dynamics and global connectivity cause suboptimal national decisions on control and prevention to have regional and even global repercussions.

Given the mismatch with regions where most expertise on infectious disease dynamics is concentrated, it is important to empower local scientists and policy-makers, in regions where the burden of disease is heaviest, about the problems facing their own countries and the consequences of local actions. It is essential to make expertise, data, models, statistical methods, and software widely available by open access. There are several initiatives (e.g., Thiswormyworld.org, Garkiproject.nd.edu, EDENextdata.com, the Humanitarian Data Exchange (HDX), and the Malaria Atlas Project), but more needs to be done. Modeling tools and software for data analysis are beginning to become open source, such that findings can be replicated, additional scenarios can be evaluated, and others can incorporate methods for data analysis or simulation. Ultimately, sharing models guarantees more reproducible results, while maximizing model transparency.

Making data sets widely available is also crucial, for example, to support replication of findings and broader comparative analyses (*79*). As models become open access, so should much of the data collected by governments, international agencies, and epidemiology research groups. Two outbreaks never occur in exactly matching circumstances, even for the same infectious agent, so there is potential to study many outbreaks in parallel to gain insight into the determinants of outbreak pattern and severity. Looking forward, there is a major opportunity to design experiments, clinical trials [for example for vaccines (*80*)], and surveillance protocols to test model predictions or assumptions, and to help reduce or better target the enormous costs involved. By integrating modeling approaches throughout the full life cycle of

infectious disease policies, including economic considerations (*58*, *70*, *81*), health outcomes can be improved and scientific understanding can be advanced.

At present, the evidence provided by infectious disease models is not considered by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group (www.gradeworkinggroup.org) alongside that of conventional studies such as clinical trials. Regardless, models are essential when diverse sources of data (including GRADE-scale evidence) need to be combined and weighed to assess quality of evidence and recommendations made in health care scenarios. In many cases, the definitive trial cannot be performed, and in such circumstances models can offer insight and extract maximum value from data that are available. In recent years, uniformity of practice and quality control for models has received more attention, resulting in initial attempts to characterize good modeling practice for infectious diseases (*82*, *83*).

The optimal use of models to inform policy decisions requires a continuous dialogue between the multidisciplinary infectious disease dynamics community and decision-makers. This is increasingly understood by governments in developed countries, in nongovernmental agencies and by large funding bodies. This dialogue will help to reduce the burden from infectious diseases by providing better-informed control strategies. Mathematical models will allow us to capitalize on new data streams and lead to an ever-greater ability to generate robust insight and collectively shape successful local and global public health policy.

REFERENCES AND NOTES

1. World Health Organization, Global strategy for health for all by the year 2000 (WHO, Geneva, 1981, http://whqlibdoc.who.int/publications/9241800038.pdf).
2. World Health Organization, Mortality and global health estimates 2013 (WHO, 2013, http://apps.who.int/gho/data/node.main.686?lang=en.).
3. P. Klepac, C. J. Metcalf, A. R. McLean, K. Hampson, Towards the endgame and beyond: Complexities and challenges for the elimination of infectious diseases. *Philos. Trans. R. Soc. Lond. B* **368**, 20120137 (2013). doi: 10.1098/rstb.2012.0137; pmid: 23798686
4. P. Daszak, A. A. Cunningham, A. D. Hyatt, Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* **287**, 443–449 (2000). doi: 10.1126/science.287.5452.443; pmid: 10642539
5. K. E. Jones *et al.*, Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008). doi: 10.1038/nature06536; pmid: 18288193
6. J. O. Lloyd-Smith *et al.*, Epidemic dynamics at the human-animal interface. *Science* **326**, 1362–1367 (2009). doi: 10.1126/science.1177345; pmid: 19965751
7. F. Keesing *et al.*, Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010). doi: 10.1038/nature09575; pmid: 21124449
8. K. F. Smith *et al.*, Global rise in human infectious disease outbreaks. *J. R. Soc. Interface* **11**, 20140950 (2014). doi: 10.1098/rsif.2014.0950; pmid: 25401184
9. S. Funk, M. Salathé, V. A. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: A review. *J. R. Soc. Interface* **7**, 1247–1256 (2010). doi: 10.1098/rsif.2010.0142; pmid: 20504800
10. C. T. Bauch, A. P. Galvani, Epidemiology. Social factors in epidemiology. *Science* **342**, 47–49 (2013). doi: 10.1126/science.1244492; pmid: 24092718
11. E. M. Volz, K. Koelle, T. Bedford, Viral phylodynamics. *PLOS Comput. Biol.* **9**, e1002947 (2013). doi: 10.1371/journal.pcbi.1002947; pmid: 23555203
12. D. Bernoulli, S. Blower, An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. *Rev. Med. Virol.* **14**, 275–288 (2004). Reprinted translation by L. Bradley with review by S. Blower; original in French 1766). doi: 10.1002/rmv.443; pmid: 15334536
13. J. O. Lloyd-Smith, D. Mollison, J. E. Metcalf, P. Klepac, J. A. P. Heesterbeek, Eds., *Challenges in Modelling Infectious Disease Dynamics*. *Epidemics* **10** (special issue, March 2015).
14. WHO Ebola Response Team, Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495 (2014). doi: 10.1056/NEJMoa1411100; pmid: 25244186
15. S. K. Gire *et al.*, Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014). doi: 10.1126/science.1259657; pmid: 25214632
16. A. Camacho *et al.*, Potential for large outbreaks of Ebola virus disease. *Epidemics* **9**, 70–78 (2014). doi: 10.1016/j.epidem.2014.09.003; pmid: 25480136
17. A. Pandey *et al.*, Strategies for containing Ebola in West Africa. *Science* **346**, 991–995 (2014). doi: 10.1126/science.1260612; pmid: 25414312
18. C. J. M. Whitty *et al.*, Infectious disease: Tough choices to reduce Ebola transmission. *Nature* **515**, 192–194 (2014). doi: 10.1038/515192a; pmid: 25391946
19. S. Blumberg, J. O. Lloyd-Smith, Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLOS Comput. Biol.* **9**, e1002993 (2013). doi: 10.1371/journal.pcbi.1002993; pmid: 23658504
20. S. Cauchemez *et al.*, Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: Application to the emergence of US swine origin influenza A H3N2v virus. *PLOS Med.* **10**, e1001399 (2013). doi: 10.1371/journal.pmed.1001399; pmid: 23472057
21. A. Kucharski, H. Mills, A. Pinsent, C. Fraser, M. Van Kerkhove, C. A. Donnelly, S. Riley, Distinguishing between reservoir exposure and human-to-human transmission for emerging pathogens using case onset data. *PLOS Curr.* (2014). pmid: 24619563
22. O. Restif *et al.*, Model-guided fieldwork: Practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol. Lett.* **15**, 1083–1094 (2012). doi: 10.1111/j.1461-0248.2012.01836.x; pmid: 22809422
23. S. Selakovic, P. C. de Ruiter, H. Heesterbeek, Infectious disease agents mediate interaction in food webs and ecosystems. *Proc. Biol. Sci.* **281**, 20132709 (2014). doi: 10.1098/rspb.2013.2709; pmid: 24403336
24. R. K. Plowright *et al.*, Ecological dynamics of emerging bat virus spillover. *Proc. Biol. Sci.* **282**, 20142124 (2015). doi: 10.1098/rspb.2014.2124; pmid: 25392474
25. M. Cotten *et al.*, Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *MBio* **5**, e01062 (2014). doi: 10.1128/mBio.01062-13; pmid: 24549846
26. K. M. Pepin, S. Lass, J. R. Pulliam, A. F. Read, J. O. Lloyd-Smith, Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat. Rev. Microbiol.* **8**, 802–813 (2010). doi: 10.1038/nrmicro2440; pmid: 20938453
27. B. T. Grenfell *et al.*, Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004). doi: 10.1126/science.1090727; pmid: 14726583
28. A. Rambaut *et al.*, The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008). doi: 10.1038/nature06945; pmid: 18418375
29. J. L. Gardy *et al.*, Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* **364**, 730–739 (2011). doi: 10.1056/NEJMoa1003176; pmid: 21345102
30. H. W. Stokes, M. R. Gillings, Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* **35**, 790–819 (2011). doi: 10.1111/j.1574-6976.2011.00273.x; pmid: 21517914
31. B. F. Koel *et al.*, Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976–979 (2013). doi: 10.1126/science.1244730; pmid: 24264991
32. B. Vrancken *et al.*, The genealogical population dynamics of HIV-1 in a large transmission chain: Bridging within and among host evolutionary rates. *PLOS Comput. Biol.* **10**, e1003505 (2014). pmid: 24699231
33. D. L. Ayres *et al.*, BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012). doi: 10.1093/sysbio/syr100; pmid: 21963610
34. S. Telfer *et al.*, Species interactions in a parasite community drive infection risk in a wildlife population. *Science* **330**, 243–246 (2010). doi: 10.1126/science.1190333; pmid: 20929776
35. I. M. Cattadori *et al.*, Infections do not predict shedding in co-infections with two helminths from a natural system. *Ecology* **95**, 1684–1692 (2014). doi: 10.1890/13-1538.1; pmid: 25039232
36. R. Anderson, J. Truscott, T. D. Hollingsworth, The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths. *Phil. Trans. R. Soc. B* **369**, 20130435 (2014). doi: 10.1098/rstb.2013.0435; pmid: 24821921
37. J. Lessler *et al.*, Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLOS Pathog.* **8**, e1002802 (2012). doi: 10.1371/journal.ppat.1002802; pmid: 22829765
38. M. A. Miller, C. Viboud, M. Balinska, L. Simonsen, The signature features of influenza pandemics—implications for policy. *N. Engl. J. Med.* **360**, 2595–2598 (2009). doi: 10.1056/NEJMp0903906; pmid: 19423872
39. J. S. Lavine, A. A. King, O. N. Bjørnstad, Natural immune boosting in pertussis dynamics and the potential for long-term vaccine failure. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7259–7264 (2011). doi: 10.1073/pnas.1014394108;pmid:21422281
40. S. Cobey, M. Lipsitch, Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science* **335**, 1376–1380 (2012). doi: 10.1126/science.1215947; pmid: 22383809
41. S. Gupta, R. M. Anderson, R. M. May, Networks of sexual contacts: Implications for the pattern of spread of HIV. *AIDS* **3**, 807–817 (1989). doi: 10.1097/00002030-198912000-00005; pmid: 2517202
42. L. Danon *et al.*, Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* **2011**, 284909 (2011). doi: 10.1155/2011/284909; pmid: 21437001
43. S. Bansal, B. T. Grenfell, L. A. Meyers, When individual behaviour matters: Homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891 (2007). doi: 10.1098/rsif.2007.1100; pmid: 17640863
44. S. Bansal, J. Read, B. Pourbohloul, L. A. Meyers, The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dyn.* **4**, 478–489 (2010). doi: 10.1080/17513758.2010.503376; pmid: 22877143
45. S. Funk, E. Gilad, C. Watkins, V. A. A. Jansen, The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6872–6877 (2009). doi: 10.1073/pnas.0810762106; pmid: 19332788
46. T. Oraby, V. Thampi, C. T. Bauch, The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc. Biol. Sci.* **281**, 20133172 (2014). doi: 10.1098/rspb.2013.3172; pmid: 24523276
47. D. Centola, The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010). pmid: 20813952
48. M. Salathé, S. Khandelwal, Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLOS Comput. Biol.* **7**, e1002199 (2011). doi: 10.1371/journal.pcbi.1002199; pmid: 22022249
49. M. J. Ferrari *et al.*, The dynamics of measles in sub-Saharan Africa. *Nature* **451**, 679–684 (2008). doi: 10.1038/nature06509; pmid: 18256664
50. T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, Will travel restrictions control the international spread of pandemic influenza? *Nat. Med.* **12**, 497–499 (2006). doi: 10.1038/nm0506-497; pmid: 16675989
51. T. A. Perkins, T. W. Scott, A. Le Menach, D. L. Smith, Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLOS Comput. Biol.* **9**, e1003327 (2013). pmid: 24348223
52. A. Wesolowski *et al.*, Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012). doi: 10.1126/science.1223467; pmid: 23066082
53. R. Ross, *The Prevention of Malaria* (John Murray, London, ed. **2**, 1911).
54. M. S. Bartlett, Measles periodicity and community size. *J. R. Stat. Soc. Ser A* **120**, 48–70 (1957). doi: 10.2307/ 2342553
55. R. M. Anderson, R. M. May, Directly transmitted infections diseases: Control by vaccination. *Science* **215**, 1053–1060 (1982). doi: 10.1126/science.7063839; pmid: 7063839

56. A. R. McLean, R. M. Anderson, Measles in developing countries. Part II. The predicted impact of mass vaccination. *Epidemiol. Infect.* **100**, 419–442 (1988). doi: 10.1017/S0950268800067170; pmid: 3378585
57. D. S. Saint-Victor, S. B. Omer, Vaccine refusal and the endgame: Walking the last mile first. *Phil. Trans. R. Soc. B* **368**, 20120148 (2013). doi: 10.1098/rstb.2012.0148; pmid: 23798696
58. P. Klepac, R. Laxminarayan, B. T. Grenfell, Synthesizing epidemiological and economic optima for control of immunizing infections. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14366–14370 (2011). doi: 10.1073/pnas.1101694108; pmid: 21825129
59. M. Salathé *et al.*, Digital epidemiology. *PLOS Comput. Biol.* **8**, e1002616 (2012). doi: 10.1371/journal.pcbi.1002616; pmid: 22844241
60. M. Salathé, C. C. Freifeld, S. R. Mekaru, A. F. Tomasulo, J. S. Brownstein, Influenza A (H7N9) and the importance of digital epidemiology. *N. Engl. J. Med.* **369**, 401–404 (2013). doi: 10.1056/NEJMp1307752; pmid: 23822655
61. A. F. Read, P. A. Lynch, M. B. Thomas, How to make evolution-proof insecticides for malaria control. *PLOS Biol.* **7**, e1000058 (2009). doi: 10.1371/journal.pbio.1000058; pmid: 19355786
62. N. Arinaminpathy, J. S. Lavine, B. T. Grenfell, Self-boosting vaccines and their implications for herd immunity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20154–20159 (2012). doi: 10.1073/pnas.1209683109; pmid: 23169630
63. O. Diekmann, J. A. P. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics* (Princeton Univ. Press, Princeton, 2013).
64. C. P. Jewell, G. O. Roberts, Enhancing Bayesian risk prediction for epidemics using contact tracing. *Biostatistics* **13**, 567–579 (2012). doi: 10.1093/biostatistics/kxs012; pmid: 22674466
65. E. S. Knock, P. D. O'Neill, Bayesian model choice for epidemic models with two levels of mixing. *Biostatistics* **15**, 46–59 (2014). doi: 10.1093/biostatistics/kxt023; pmid: 23887980
66. J. Dureau, K. Kalogeropoulos, M. Baguelin, Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics* **14**, 541–555 (2013). doi: 10.1093/biostatistics/kxs052; pmid: 23292757
67. A. A. King, E. L. Ionides, M. Pascual, M. J. Bouma, Inapparent infections and cholera dynamics. *Nature* **454**, 877–880 (2008). doi: 10.1038/nature07084; pmid: 18704085
68. O. Ratmann, G. Donker, A. Meijer, C. Fraser, K. Koelle, Phylodynamic inference and model assessment with approximate bayesian computation: Influenza as a case study. *PLOS Comput. Biol.* **8**, e1002835 (2012). doi: 10.1371/journal.pcbi.1002835; pmid: 23300420
69. M. Farah, P. J. Birrell, S. Conti, D. De Angelis, Bayesian Emulation and Calibration of a Dynamic Epidemic Model for A/H1N1 Influenza. *J. Am. Stat. Assoc.* **109**, 1398–1411 (2014). doi: 10.1080/01621459.2014.934453
70. M. Baguelin *et al.*, Assessing optimal target populations for influenza vaccination programmes: An evidence synthesis and modelling study. *PLOS Med.* **10**, e1001527 (2013). doi: 10.1371/journal.pmed.1001527; pmid: 24115913
71. D. De Angelis, A. M. Presanis, S. Conti, A. E. Ades, Estimation of HIV Burden through Bayesian Evidence Synthesis. *Stat. Sci.* **29**, 9–17 (2014). doi: 10.1214/13-STS428
72. J. Ginsberg *et al.*, Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1014 (2009). doi: 10.1038/nature07634; pmid: 19020500
73. N. Bharti *et al.*, Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334**, 1424–1427 (2011). doi: 10.1126/science.1210554; pmid: 22158822
74. J. P. Messina *et al.*, Global spread of dengue virus types: Mapping the 70 year history. *Trends Microbiol.* **22**, 138–146 (2014). doi: 10.1016/j.tim.2013.12.011; pmid: 24468533
75. C. Dye *et al.*, WHO and the future of disease control programmes. *Lancet* **381**, 413–418 (2013). doi: 10.1016/S0140-6736(12)61812-1; pmid: 23374479
76. C. Dye, After 2015: Infectious diseases in a new era of health and development. *Philos. Trans. R. Soc. Lond. B* **369**, 20130426 (2014). doi: 10.1098/rstb.2013.0426; pmid: 24821913
77. M. Woolhouse, J. Farrar, Policy: An intergovernmental panel on antimicrobial resistance. *Nature* **509**, 555–557 (2014). doi: 10.1038/509555a; pmid: 24877180
78. R. F. Terry, J. F. Salm Jr., C. Nannei, C. Dye, Creating a global observatory for health R&D. *Science* **345**, 1302–1304 (2014). doi: 10.1126/science.1258737; pmid: 25214621
79. P. Rohani, A. A. King, Never mind the length, feel the quality: The impact of long-term epidemiological data sets on theory, application and policy. *Trends Ecol. Evol.* **25**, 611–618 (2010). doi: 10.1016/j.tree.2010.07.010; pmid: 20800928
80. M. E. Halloran, I. M. Longini, C. J. Struchiner, *Design and Analysis of Vaccine studies* (Springer, Berlin, 2010).
81. M. L. Ndeffo Mbah *et al.*, Cost-effectiveness of a community-based intervention for reducing the transmission of *Schistosoma haematobium* and HIV in Africa. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7952–7957 (2013). doi: 10.1073/pnas.1221396110; pmid: 23589884
82. R. Pitman *et al.*, Dynamic transmission modeling: A report of the ISPOR-SMDM modeling good research practices task force-5. *Value Health* **15**, 828–834 (2012). doi: 10.1016/j.jval.2012.06.011; pmid: 22999132
83. M. E. J. Woolhouse *et al.*, Guide to good practice for quantitative veterinary epidemiology. www.qve-goodpracticeguide.org.uk/ (2011).
84. T. D. Hollingsworth, D. Klinkenberg, H. Heesterbeek, R. M. Anderson, Mitigation strategies for pandemic influenza A: Balancing conflicting policy objectives. *PLOS Comput. Biol.* **7**, e1001076 (2011). doi: 10.1371/journal.pcbi.1001076; pmid: 21347316
85. J. T. Wu, B. J. Cowling, The use of mathematical models to inform influenza pandemic preparedness and response. *Exp. Biol. Med.* **236**, 955–961 (2011). doi: 10.1258/ebm.2010.010271; pmid: 21727183
86. B. Y. Lee, A. E. Wiringa, The 2009 H1N1 influenza pandemic: A case study of how modeling can assist all stages of vaccine decision-making. *Hum. Vaccin.* **7**, 115–119 (2011). doi: 10.4161/hv.7.1.13740; pmid: 21263227
87. M. Lipsitch, L. Finelli, R. T. Heffernan, G. M. Leung, S. C. Redd; 2009 H1n1 Surveillance Group, Improving the evidence base for decision making during a pandemic: The example of 2009 influenza A/H1N1. *Biosecur. Bioterror.* **9**, 89–115 (2011). pmid: 21612363
88. T. House *et al.*, Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. *Proc. Biol. Sci.* **278**, 2753–2760 (2011). doi: 10.1098/rspb.2010.2688; pmid: 21288945
89. M. D. Van Kerkhove, N. M. Ferguson, Epidemic and intervention modelling—a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic. *Bull. World Health Organ.* **90**, 306–310 (2012). pmid: 22511828
90. M. Luksza, M. Lässig, A predictive fitness model for influenza. *Nature* **507**, 57–61 (2014). doi: 10.1038/nature13087; pmid: 24572367
91. UK Government, (**2012**, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224775/JCVI-statement-on-the-annual-influenza-vaccination-programme-25-July-2012.pdf).
92. M. A. Rose *et al.*, The epidemiological impact of childhood influenza vaccination using live-attenuated influenza vaccine (LAIV) in Germany: Predictions of a simulation study. *BMC Infect. Dis.* **14**, 40 (2014). doi: 10.1186/1471-2334-14-40; pmid: 24450996
93. HIV Modelling Consortium Treatment as Prevention Editorial Writing Group, HIV treatment as prevention: Models, data, and questions—towards evidence-based decision-making. *PLOS Med.* **9**, e1001259 (2012). doi: 10.1371/journal.pmed.1001259; pmid: 22802739
94. A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, D. D. Ho, HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996). doi: 10.1126/science.271.5255.1582; pmid: 8599114
95. N. R. Faria *et al.*, The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014). doi: 10.1126/science.1256739; pmid: 25278604
96. T. D. Hollingsworth, R. M. Anderson, C. Fraser, HIV-1 transmission, by stage of infection. *J. Infect. Dis.* **198**, 687–693 (2008). doi: 10.1086/590501; pmid: 18662132
97. D. Bezemer *et al.*, A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. *AIDS* **22**, 1071–1077 (2008). doi: 10.1097/QAD.0b013e3282fd167c; pmid: 18520351
98. J. S. Montaner *et al.*, The case for expanding access to highly active antiretroviral therapy to curb the growth of the HIV epidemic. *Lancet* **368**, 531–536 (2006). doi: 10.1016/S0140-6736(06)69162-9; pmid: 16890841
99. R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, B. G. Williams, Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. *Lancet* **373**, 48–57 (2009). doi: 10.1016/S0140-6736(08)61697-9; pmid: 19038438
100. M. S. Cohen *et al.*, Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med.* **365**, 493–505 (2011). doi: 10.1056/NEJMoa1105243; pmid: 21767103
101. S. E. Bellan *et al.*, Extra-couple HIV transmission in sub-Saharan Africa: A mathematical modelling study of survey data. *Lancet* **381**, 1561–1569 (2013). doi: 10.1016/S0140-6736(12)61960-6; pmid: 23391466
102. F. Tanser, T. Bärnighausen, E. Grapsa, J. Zaidi, M. L. Newell, High coverage of ART associated with decline in risk of HIV acquisition in rural KwaZulu-Natal, South Africa. *Science* **339**, 966–971 (2013). doi: 10.1126/science.1228160; pmid: 23430656
103. S. H. Vermund, S. J. Fidler, H. Ayles, N. Beyers, R. J. Hayes, Can combination prevention strategies reduce HIV transmission in generalized epidemic settings in Africa? The HPTN 071 (PopART) study plan in South Africa and Zambia. *J. Acquir. Immune Defic. Syndr.* **63** (suppl. 2), S221–S227 (2013). doi: 10.1097/QAI.0b013e318299c3f4; pmid: 23764639
104. M. K. Smith, K. A. Powers, K. E. Muessig, W. C. Miller, M. S. Cohen, HIV treatment as prevention: The utility and limitations of ecological observation. *PLOS Med.* **9**, e1001260 (2012). doi: 10.1371/journal.pmed.1001260; pmid: 22802740
105. M. S. Cohen *et al.*, Antiretroviral treatment of HIV-1 prevents transmission of HIV-1: Where do we go from here? *Lancet* **382**, 1515–1524 (2013). doi: 10.1016/S0140-6736(13)61998-4; pmid: 24152938
106. G. MacDonald, The analysis of equilibrium in malaria. *Trop. Dis. Bull.* **49**, 813–829 (1952). pmid: 12995455
107. K. Dietz, L. Molineaux, A. Thomas, A malaria model tested in the African savannah. *Bull. World Health Organ.* **50**, 347–357 (1974). pmid: 4613512
108. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1992).
109. A. S. Perelson, R. M. Ribeiro, Modeling the within-host dynamics of HIV infection. *BMC Biol.* **11**, 96 (2013). doi: 10.1186/1741-7007-11-96; pmid: 24020860
110. C. A. Donnelly, N. M. Ferguson, **2000**, *Statistical Aspects of BSE and vCJD: Models for Epidemics* (Chapman & Hall/CRC, Boca Raton, FL, 2000).
111. M. J. Keeling, Models of foot-and-mouth disease. *Proc. Biol. Sci.* **272**, 1195–1202 (2005). doi: 10.1098/rspb.2004.3046; pmid: 16024382
112. R. M. Anderson *et al.*, Epidemiology, transmission dynamics and control of SARS: The 2002-2003 epidemic. *Phil. Trans. R. Soc. B* **359**, 1091–1105 (2004). doi: 10.1098/rstb.2004.1490; pmid: 15306395
113. S. Blumberg, J. O. Lloyd-Smith, Comparing methods for estimating R_0 from the size distribution of subcritical transmission chains. *Epidemics* **5**, 131–145 (2013). doi: 10.1016/j.epidem.2013.05.002; pmid: 24021520
114. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005). doi: 10.1038/nature04153; pmid: 16292310
115. L. Matthews *et al.*, Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 547–552 (2006). doi: 10.1073/pnas.0503776103; pmid: 16407143
116. N. M. Ferguson, A. P. Galvani, R. M. Bush, Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433 (2003). doi: 10.1038/nature01509; pmid: 12660783
117. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006). doi: 10.1126/science.1132745; pmid: 17185596
118. M. Recker, O. G. Pybus, S. Nee, S. Gupta, The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7711–7716 (2007). doi: 10.1073/pnas.0702154104; pmid: 17460037
119. S. Funk, H. Nishiura, H. Heesterbeek, W. J. Edmunds, F. Checchi, Identifying transmission cycles at the human-animal interface: The role of animal reservoirs in maintaining gambiense human african trypanosomiasis. *PLOS Comput. Biol.* **9**, e1002855 (2013). doi: 10.1371/journal.pcbi.1002855; pmid: 23341760
120. R. C. Reiner Jr. *et al.*, A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *J. R. Soc. Interface* **10**, 20120921 (2013). doi: 10.1098/rsif.2012.0921; pmid: 23407571
121. E. M. Stuckey, T. A. Smith, N. Chitnis, Estimating malaria transmission through mathematical models. *Trends Parasitol.* **29**, 477–482 (2013). doi: 10.1016/j.pt.2013.08.001; pmid: 24001452
122. N. G. Reich *et al.*, Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity.

J. R. Soc. Interface **10**, 20130414 (2013). doi: 10.1098/rsif.2013.0414; pmid: 23825116
123. B. S. Cooper *et al.*, Methicillin-resistant *Staphylococcus aureus* in hospitals and the community: Stealth dynamics and control catastrophes. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10223–10228 (2004). doi: 10.1073/pnas.0401324101; pmid: 15220470
124. C. Dye, B. G. Williams, The population dynamics and control of tuberculosis. *Science* **328**, 856–861 (2010). doi: 10.1126/science.1185449; pmid: 20466923
125. A. F. Read, T. Day, S. Huijben, The evolution of drug resistance and the curious orthodoxy of aggressive chemotherapy. *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 2), 10871–10877 (2011). doi: 10.1073/pnas.1100299108; pmid: 21690376
126. S. Cobey, M. Lipsitch, Pathogen diversity and hidden regimes of apparent competition. *Am. Nat.* **181**, 12–24 (2013). doi: 10.1086/668598; pmid: 23234842
127. C. J. E. Metcalf *et al.*, Structured models of infectious disease: Inference with discrete data. *Theor. Popul. Biol.* **82**, 275–282 (2012). doi: 10.1016/j.tpb.2011.12.001; pmid: 22178687
128. J. Lessler, C. J. E. Metcalf, Balancing evidence and uncertainty when considering rubella vaccine introduction. *PLOS ONE* **8**, e67639 (2013). doi: 10.1371/journal.pone.0067639; pmid: 23861777
129. R. Anderson, T. D. Hollingsworth, J. Truscott, S. Brooker, Optimisation of mass chemotherapy to control soil-transmitted helminth infection. *Lancet* **379**, 289–290 (2012). doi: 10.1016/S0140-6736(12)60120-2; pmid: 22284644
130. C. J. E. Metcalf *et al.*, Implications of spatially heterogeneous vaccination coverage for the risk of congenital rubella syndrome in South Africa. *J. R. Soc. Interface* **10**, 20120756 (2013). doi: 10.1098/rsif.2012.0756; pmid: 23152104
131. T. Panagiotopoulos, I. Antoniadou, E. Valassi-Adam, A. Berger, Increase in congenital rubella occurrence after immunisation in Greece: Retrospective survey and systematic review. *BMJ* **319**, 1462–1467 (1999). doi: 10.1136/bmj.319.7223.1462; pmid: 10582926
132. C. H. Chan, C. J. McCabe, D. N. Fisman, Core groups, antimicrobial resistance and rebound in gonorrhoea in North America. *Sex. Transm. Infect.* **88**, 200–204 (2012). doi: 10.1136/sextrans-2011-050049; pmid: 22169277
133. Y. Nagao, K. Koelle, Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2238–2243 (2008). doi: 10.1073/pnas.0709029105; pmid: 18250338

ACKNOWLEDGMENTS

H.H. conceived and wrote the paper; R.M.A., V.A., S.B., D.D.A., C.D., K.T.D.E., W.J.E., S.D.W.F., S.F., T.D.H., T.H., V.I., P.K., J.L., J.O.L.-S., C.J.E.M., D.M., J.R.C.P., L.P., M.G.R., and C.V. provided text and edited the manuscript; J.L., C.J.E.M. and T.D.H. produced figures; the Isaac Newton Institute IDD Collaboration jointly produced and discussed ideas for the outline and content (all of the above plus N.A., F.B., T.B., J.G., B.G., A.L.L., A.M., P.O.N., C.P., S.R., G.S.T., P.T., and J.W.). We gratefully acknowledge help by K. Koelle and D. Fisman in producing adapted versions of their figures for Fig. 2 (panels C, D, and E). This paper was conceived and developed at a program on Infectious Disease Dynamics at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, 19 August to 13 September 2013 and 19 May to 6 June 2014 (www.newton.ac.uk). We gratefully acknowledge financial and infrastructural support from the Isaac Newton Institute for Mathematical Sciences which is fundamental to the success of this program. We are also grateful for the financial support the program received from the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science and Technology Directorate, U.S. Department of Homeland Security, and the Fogarty International Center, NIH. R.M.A. is a non–executive director of GlaxoSmithKline. C.D. acts as an advisor for the Wellcome Trust, for which he receives financial remuneration.

10.1126/science.aaa4339

MICROBIOLOGY

NEWS

Computer scan uncovers 100,000 new viruses

Clues to future outbreaks may be hidden in existing genomic databases

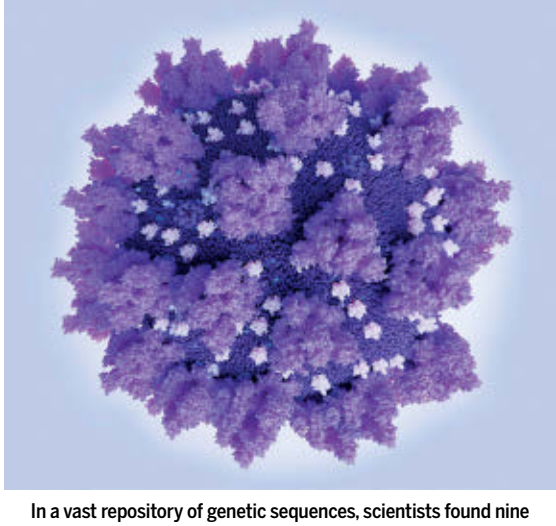
By Elizabeth Pennisi

It took just one virus to cripple the world's economy and kill millions of people; yet virologists estimate that trillions of still-unknown viruses exist, many of which might be lethal or have the potential to spark the next pandemic. Now, they have a new—and very long—list of possible suspects to interrogate. By sifting through unprecedented amounts of existing genomic data, scientists have uncovered more than 100,000 novel viruses, including nine coronaviruses and more than 300 related to the hepatitis Delta virus, which can cause liver failure.

"It's a foundational piece of work," says J. Rodney Brister, a bioinformatician at the National Library of Medicine. The study, published last week in *Nature*, expands the number of known viruses that use RNA instead of DNA for their genes by an order of magnitude. It "demonstrates our outrageous lack of knowledge about this group of organisms," says disease ecologist Peter Daszak, president of the EcoHealth Alliance, a nonprofit research group in New York City that is raising money to launch a global survey of viruses.

Scientists predict the study will also help launch so-called petabyte genomics—the analyses of previously unfathomable quantities of DNA and RNA data. (One petabyte is 10^{15} bytes.) That wasn't exactly what computational biologist Artem Babaian had in mind when he came up with the project while in between jobs in early 2020. Instead, he was simply curious about how many coronaviruses—aside from the virus that had just launched the COVID-19 pandemic—could be found in sequences in existing genomic databases.

So, he and independent supercomputing expert Jeff Taylor scoured cloud-based genomic data that had been deposited to a global sequence database and uploaded by the U.S. National Institutes of Health. As of now, the database contains 16 petabytes of archived sequences, which come from genetic surveys of everything from fugu fish, the risky Japanese delicacy, to farm soils to human guts. (A database with a 5-megabase digital photo of every person in the United States would take up about the same amount of space.) The sequences also capture the genomes of viruses infecting different organisms in samples, but the viruses usually go undetected.



In a vast repository of genetic sequences, scientists found nine unknown coronaviruses, relatives of SARS-CoV-2 (computer model).

IMAGE: JUAN GAERTNER/SCIENCE SOURCE

To sift through the reams of data, Babaian and Taylor devised a set of computer search tools specialized for cloud-based data. With the help of several bioinformaticians, some whom became collaborators on the project, they tweaked the new software to make their analysis "way faster than anyone thought possible," recalls Babaian, who is now at the University of Cambridge.

They soon expanded the viral hunt beyond coronaviruses and looked at all the data in the cloud. Babaian and his colleagues' programs hunted among the cloud's sequences for matches to the central core of the gene for RNA-dependent RNA polymerase, which is key to the replication of all RNA viruses. Such viruses include not only coronaviruses, but also those that cause flu, polio, measles, and hepatitis.

Babaian's approach was fast enough to work through 1 million data sets a day—at a com-

puting cost of less than 1 cent per data set. "It's an impressive engineering feat," says C. Titus Brown, a bioinformatician at the University of California, Davis. When the researchers were finally finished, they had uncovered the partial genomes of almost 132,000 RNA viruses.

The group's new database doesn't have the complete sequence of each new virus— in many cases, there's just the gene for the core enzyme. But researchers can use even partial sequences to build family trees that reveal how different viruses are related. In some cases, they can also use the database to find out where around the world a particular virus was found—and what type of host it was in. And some of the discovered viruses could help researchers better understand how human pathogens arise, Brown says, or improve diagnostic tests for infections. Finally, when a new virus is isolated from a sick patient, a scan of the genomic database could show whether it was already present elsewhere. "We have turned this [database] into a giant virus surveillance network," Babaian says.

Some findings were unexpected, including new coronaviruses in the well-studied fugu fish and in the axolotl, an amphibian that is a common lab organism. In a few cases, researchers could piece together whole genomes for the viral finds. And in some aquatic animals, those sequences suggested their novel coronavirus genomes are spread across two separate RNA molecules, not the usual single strand, Babaian and his colleagues report.

Babaian's team also came across evidence of more than 250 giant bacteriophages—viruses that infect bacteria—that resemble ones already known in algae. These "huge phages" were detected in sequences from vastly different organisms. One group of huge phages was found in a per-son in Bangladesh and also in cats and dogs in the United Kingdom, for example. These viruses are big enough to carry genes between different hosts species, suggesting they might provide a new source of genetic changes, Babaian notes. That's the way it is with viruses, Daszak says. "Every time we start digging, we get surprises."

To make sure others can take advantage of the work, Babaian's team has created a public repository of the tools it developed, along with the results. The amount of cloud-based, publicly available DNA sequences is expanding exponentially; if he did the same analysis next year, Babaian expects he would find hundreds of thousands more RNA viruses. "By the end of decade, I want to identify over 100 million." ■

HIV

A highly virulent variant of HIV-1 circulating in the Netherlands

Chris Wymant[1]*, Daniela Bezemer[2], François Blanquart[3,4], Luca Ferretti[1], Astrid Gall[5], Matthew Hall[1], Tanya Golubchik[1], Margreet Bakker[6], Swee Hoe Ong[7], Lele Zhao[1], David Bonsall[1,8], Mariateresa de Cesare[8], George MacIntyre-Cockett[1,8], Lucie Abeler-Dörner[1], Jan Albert[9,10], Norbert Bannert[11], Jacques Fellay[12,13,14], M. Kate Grabowski[15], Barbara Gunsenheimer-Bartmeyer[16], Huldrych F. Günthard[17,18], Pia Kivelä[19], Roger D. Kouyos[17,18], Oliver Laeyendecker[20], Laurence Meyer[21], Kholoud Porter[22], Matti Ristola[19], Ard van Sighem[2], Ben Berkhout[6], Paul Kellam[23,24], Marion Cornelissen[6,25], Peter Reiss[2,26], Christophe Fraser[1,8]*, the Netherlands ATHENA HIV Observational Cohort†, the BEEHIVE Collaboration†

We discovered a highly virulent variant of subtype-B HIV-1 in the Netherlands. One hundred nine individuals with this variant had a 0.54 to 0.74 $\log_{10}$ increase (i.e., a ~3.5-fold to 5.5-fold increase) in viral load compared with, and exhibited CD4 cell decline twice as fast as, 6604 individuals with other subtype-B strains. Without treatment, advanced HIV—CD4 cell counts below 350 cells per cubic millimeter, with long-term clinical consequences—is expected to be reached, on average, 9 months after diagnosis for individuals in their thirties with this variant. Age, sex, suspected mode of transmission, and place of birth for the aforementioned 109 individuals were typical for HIV-positive people in the Netherlands, which suggests that the increased virulence is attributable to the viral strain. Genetic sequence analysis suggests that this variant arose in the 1990s from de novo mutation, not recombination, with increased transmissibility and an unfamiliar molecular mechanism of virulence.

The risk posed by viruses evolving to greater virulence—i.e., causing greater damage to their hosts—has been extensively studied in theoretical work despite few population-level examples (*1*–*3*). The most notable recent example is the B.1.617.2 lineage (Delta variant) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), for which an increased probability of death has been reported (*4*–*6*), as well as increased transmissibility (*7*, *8*). RNA viruses have long been a particular concern, as their error-prone replication results in the greatest known rate of mutation—and thus high potential for adaptation. Greater virulence could benefit a virus if it is not outweighed by reduced opportunity for transmission. These antagonistic selection pressures may result in an intermediate level of virulence being optimal for viral fitness, as observed for HIV (*9*). Concrete examples of such evolution in action, however, have been elusive. Continued monitoring of HIV virulence is important for global health: 38 million people currently live with the

Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [2]Stichting HIV Monitoring, Amsterdam, Netherlands. [3]Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France. [4]IAME, UMR 1137, INSERM, Université de Paris, Paris, France. [5]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. [6]Laboratory of Experimental Virology, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. [7]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK. [8]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [9]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden. [10]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden. [11]Division for HIV and Other Retroviruses, Department of Infectious Diseases, Robert Koch Institute, Berlin, Germany. [12]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [13]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [14]Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland. [15]Department of Pathology, John Hopkins University, Baltimore, MD, USA. [16]Department of Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany. [17]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland. [18]Institute of Medical Virology, University of Zurich, Zurich, Switzerland. [19]Department of Infectious Diseases, Helsinki University Hospital, Helsinki, Finland. [20]Division of Intramural Research, NIAID, NIH, Baltimore, MD, USA. [21]INSERM CESP U1018, Université Paris Saclay, APHP, Service de Santé Publique, Hôpital de Bicêtre, Le Kremlin-Bicêtre, France. Institute for Global Health, University College London, London, UK. [23]Kymab Ltd., Cambridge, UK. [24]Department of Infectious Diseases, Faculty of Medicine, Imperial College London, London, UK. [25]Molecular Diagnostic Unit, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. Department of Global Health, Amsterdam University Medical Centers, University of Amsterdam and Amsterdam Institute for Global Health and Development, Amsterdam, Netherlands.
***Corresponding author. Email: chris.wymant@bdi.ox.ac.uk (C.W.); christophe.fraser@bdi.ox.ac.uk (C.F.)** †Contributors and affiliations are listed in the supplementary materials.

virus, and it has caused an estimated 33 million deaths (www.unaids.org).

The main (M) group of HIV-1, responsible for the global pandemic, first emerged around 1920 in the area of what is now Kinshasa, Democratic Republic of the Congo (*10*), and had diversified into subtypes by 1960 (*11*). The subtypes, and the most common circulating recombinant forms (CRFs) between the subtypes, took different routes for global spread, establishing strong associations with geography (*12*), ethnicity, and mode of transmission. Differences in virulence between subtypes and CRFs have been reported, though it is challenging to disentangle genotypic effects on virulence from confounding effects while retaining large sample sizes, given the strong associations between viral, host, and epidemiological factors (*13*). The co-receptor used for cell entry has long been understood to affect virulence (*14*, *15*), and this has been proposed as a mechanism that underlies differences in virulence between subtypes and CRFs (*13*), as well as one reported difference within a CRF (*16*).

HIV-1 virulence is most commonly measured by viral loads (the concentration of viral particles in blood plasma) and CD4 counts (the concentration of CD4+ T cells in peripheral blood, which tracks immune system damage by the virus). Successful treatment with antiretroviral drugs suppresses viral load and interrupts the decline in CD4 counts that would otherwise lead to AIDS. Both viral load and rate of CD4 cell decline are heritable properties—that is, these properties are causally affected by viral genetics, leading to correlation between an individual and whomever they infect (*17*–*21*). It has therefore been expected that viral load and CD4 cell decline could change with the emergence of a new viral variant. We substantiate that expectation with empirical evidence by reporting a subtype-B variant of HIV-1 with exceptionally high virulence that has been circulating within the Netherlands during the past two decades.

Discovery of the highly virulent variant

Within an ongoing study (the BEEHIVE project; www.beehive.ox.ac.uk), we identified a group of 17 individuals with a distinct subtype-B viral variant, whose viral loads in the set-point window of infection (6 to 24 months after a positive test obtained early in the course of infection) were highly elevated (Table 1, middle column). BEEHIVE is a study of individuals enrolled in eight cohorts across Europe and Uganda, who were selected because they have well-characterized dates of infection and samples available from early infection, for whom whole viral genomes were sequenced. The 17 individuals with the distinct viral variant comprised 15 participants in the ATHENA study in the Netherlands, 1 from Switzerland, and 1 from Belgium. See materials and methods for details on the initial discovery.

Replication of the discovery in Dutch ATHENA data

To replicate the finding and to investigate this viral variant in more detail, we then analyzed data from 6706 participants in ATHENA with subtype-B infections (expanding on the subset of 521 participants in ATHENA who were eligible for inclusion in BEEHIVE). We found 92 additional individuals infected with the viral variant, bringing the total to 109 such individuals in either dataset. When replicating the BEEHIVE test with the ATHENA data (Table 1, right column), we again observed a large rise in viral load in individuals with this viral variant: an increase of 0.54 $\log_{10}$ viral copies/ml (i.e., a ~3.5-fold increase). The effect size was the same in a linear model including age at diagnosis and sex as covariates, and persisted in newly diagnosed individuals over time (Fig. 1A). Henceforth, for brevity, we refer to this viral variant as the "VB variant" (for virulent subtype B), to individuals infected with this variant as "VB individuals," and to individuals infected with a different strain of HIV as "non-VB individuals."

Search for closely related viruses

To test whether the variant was more widely disseminated, we searched publicly available databases for similar HIV viral genotypes. All results had <95% sequence similarity to a representative viral sequence for the variant. Of the 17 VB individuals originally found in BEEHIVE, one was from the Swiss HIV Cohort Study (*22*) (SHCS). By examining previously published data (*23*), we found that three other individuals from the SHCS were closely related (a phylogenetic distance below 2.5%). The high coverage of the Swiss HIV Cohort [including 89% of reported new infections from 2009 through 2018, with ~65% of the cohort sequenced (*24*)] makes it unlikely that many more VB individuals in Switzerland were undetected. Data to assess viral load or CD4 cell decline for these three individuals were not available, owing to early initiation of treatment.

More-rapid CD4 cell decline

At the time of diagnosis, CD4 counts for VB individuals were already lower than for non-VB individuals by 73 cells/mm^3 [95% confidence interval (CI): 12 to 134]. These counts subsequently declined faster, by a further 49 cells/mm^3 per year (CI: 20 to 79), in addition to the decline for comparable non-VB individuals [49 cells/mm^3 per year (CI: 46 to 51) for men diagnosed at the age of 30 to 39 years]. The VB variant is therefore associated with a doubling in the rate of CD4 cell decline. These values are averages estimated by using a linear mixed model adjusted for sex and age at diagnosis. Figure 1B illustrates the CD4 count decline that would be expected if disease progression were to continue linearly in the absence of treatment. Initiating treatment at a CD4 count of 350 cells/mm^3, instead of immediately, was previously shown to substantially increase the subsequent hazard for serious adverse events (*25*). As seen in Fig. 1B, this stage of CD4 cell decline is expected to be reached in 9 months (CI: 2 to 17) from the time of diagnosis for VB individuals, as opposed to 36 months (CI: 33 to 39) for non-VB individuals, in males diagnosed at the age of 30 to 39 years. It is reached even more quickly in older age groups, for which we found progressively lower CD4 counts at time of diagnosis (table S1). At a CD4 count of 200 cells/mm^3, there is a high risk of immediate AIDS-related complications; without treatment this stage of decline would be reached, on average, between 2 and 3 years after diagnosis for VB individuals and between 6 and 7 years after diagnosis for comparable non-VB individuals [the latter being similar to previous reports in Europe (*26*)].

The effect of the VB variant on CD4 cell decline remained after we adjusted for the effect of higher viral load. With this adjustment, VB individuals have a CD4 count at diagnosis as would be expected given their high viral loads, but their subsequent decline in CD4 counts is again twice as fast as for as comparable non-VB individuals with high viral loads—their rate of decline is accelerated by 44 cells/mm^3 per year (CI: 16 to 72). Comparison of this additional decline with that expected from a +1 increase in $\log_{10}$ viral load, 15 cells/mm^3 per year (CI: 11 to 18), shows that the variant's effect on CD4 count decline is equivalent to that expected from a +3.0 increase in $\log_{10}$ viral load. The same analysis of measurements of CD4 percentages (the percentage of all T cells that express CD4) showed that these also declined twice as fast for VB individuals, and again this doubling in speed of decline remained when we adjusted for the higher viral load of the variant (table S2 and fig. S1).

No difference in CD4 cells after treatment, or in mortality

Measurements of treatment success include CD4 cell recovery and mortality. CD4 counts and percentages after treatment initiation were similar for VB and non-VB individuals, as measured with both linear mixed modeling of the CD4 dynamics (tables S3 and S4 and fig. S2) and an individual-matching procedure. The hazard for death (from any cause) was also similar: VB individuals had a relative hazard of 1.4 (CI: 0.7 to 2.8, $P = 0.35$, Cox proportional hazards model). Our study had statistical power to detect only very large differences in mortality, as reflected

in the wide CI for relative hazard for death and shown in Fig. 1C. VB individuals had similar CD4 counts and mortality after treatment despite a faster CD4 cell decline before treatment; this could be explained by their tendency to start treatment sooner after diagnosis (fig. S3). For example, although the probability of having started treatment was estimated to be similar 6 months after diagnosis [42% (CI: 41 to 44%) for non-VB individuals compared with 46% (CI: 35 to 54%) for VB individuals], it was different 2 years after diagnosis [65% (CI: 64 to 67%) for non-VB individuals and 93% (CI: 85 to 96%) for VB individuals]. Had VB individuals not started treatment earlier than others, lower CD4 counts at treatment initiation would have been expected, potentially causing increased morbidity and mortality (*25*). This information could be relevant if VB or variants like it are found in settings with less widespread availability of HIV care.

Characteristics of individuals infected with the VB variant

VB individuals were mostly (82%) men who have sex with men, similar to non-VB individuals (76%). Age at diagnosis was also similar for VB and non-VB individuals (fig. S4). Neither ethnicity nor host genotype data were available, but the place of birth was mostly recorded as Western Europe for both groups (71% for non-VB individuals, 86% for VB individuals). VB individuals were present in all regions of the Netherlands, but with a different distribution relative to that of non-VB individuals (N = 102 versus 6604 individuals, $P < 10^{-7}$, simulated Fisher's exact test): VB individuals were more common in the south (25% of VB individuals versus 6% of non-VB individuals) and less common in Amsterdam (20% versus 51%), as shown in table S5. Table S6 lists the hospitals included in each region. The average time from infection to diagnosis, for men who have sex with men in this cohort diagnosed in the late 2000s, was previously estimated to be 3.6 years (CI: 3.3 to 4.0) (*27*).

Genotype of the VB variant

Sequence data from the BEEHIVE project are whole-genome data, providing the 17 whole genomes available for the variant; sequence data from ATHENA are partial *pol* gene data only, available for the additional 92 VB individuals. We subtyped the 17 whole genomes for the variant as pure subtype B [with 100% support from two concordant methods (*28*, *29*)], like most HIV-1 in the Netherlands. We predicted co-receptor usage from the 17 whole genomes using two concordant methods (*30*, *31*): one was likely CXCR4-tropic; the other 16 were likely CCR5-tropic. Only one drug-resistance mutation was common for the VB variant: $Met^{41}\rightarrow Leu$ (M41L), present in 91 of 109 partial *pol* gene sequences. Without other linked resistance mutations, M41L causes only low-level resistance to zidovudine (*32*, *33*). Two of the whole genomes were found to be recombinants between the VB variant and another subtype-B cluster in ATHENA (containing a small amount of sequence from the latter) and were excluded from subsequent sequence analysis. Among whole genomes in BEEHIVE and all whole genomes in the Los Alamos National Laboratory HIV Database (www.hiv.lanl.gov), none appeared to be a candidate for a "recombination parent" of the VB variant—i.e., the many mutations that distinguish the VB variant from any other known virus appear to have arisen de novo, not through recombination.

We compared the consensus sequence for the VB variant with the consensus of all Dutch subtype-B sequences in BEEHIVE, at both the amino acid and the nucleotide level: There were 250 amino acid changes and 509 nucleotide changes, as well as insertions and deletions. These alignments are included as data S1, and the amino acid alignment is illustrated in fig. S5. The distribution of nucleotide changes over the genome is in line with expectations (for example, fewer in the conserved *pol* gene region and more in the variable *env* gene region; see fig. S6). The VB-variant genotype is thus characterized by many mutations spread through the genome, meaning that a single genetic cause for the enhanced virulence cannot be determined from the current data.

We conducted descriptive analyses of the mutations that distinguish the VB variant from the Dutch subtype-B consensus. All of the amino acid–level changes are listed in data S2 with annotations. Of the observed amino acid substitutions, 30 were previously shown to be positively associated with escape from cytotoxic T lymphocyte (CTL) response for at least one human leukocyte antigen type, and 13 were shown to be negatively associated

Table 1. Comparison of viral loads between individuals infected with the VB viral variant and other individuals. When analyzing the viral loads of individuals in the ATHENA study, we first excluded individuals who were in BEEHIVE, so that the test would be independent of the initial finding within the BEEHIVE study. After our statistical tests of viral load, we did not exclude BEEHIVE individuals from the ATHENA data for subsequent analyses. *N*, number of individuals after those without viral load measurements before treatment were excluded; IQR, interquartile range.

Test	Discovery [BEEHIVE dataset (Europe)]	Replication [ATHENA dataset (Netherlands), excluding overlap with BEEHIVE]
Viral load measurements compared	Set-point viral loads for N = 15 VB individuals and N = 2446 individuals with any other HIV-1 strain	Mean pretreatment viral loads for N = 91 VB individuals and N = 5272 individuals with any other subtype-B HIV-1 strain
Mean and IQR of viral load in non-VB individuals, in $\log_{10}$ copies per milliliter	5.10 (IQR: 4.69 to 5.58)	4.79 (IQR: 4.34 to 5.27)
Mean and IQR of viral load in VB individuals, in $\log_{10}$ copies per milliliter	5.84 (IQR: 5.57 to 6.09)	5.33 (IQR: 4.94 to 5.75)
Viral load increase in VB individuals	0.74 $\log_{10}$ viral copies/ml	0.54 $\log_{10}$ viral copies/ml
P value for increase	5×10^{-6} (two-tailed t test, significant at a level of 5×10^{-5} when Bonferroni-corrected for performing 50 such tests)	1×10^{-12} (one-tailed t test)

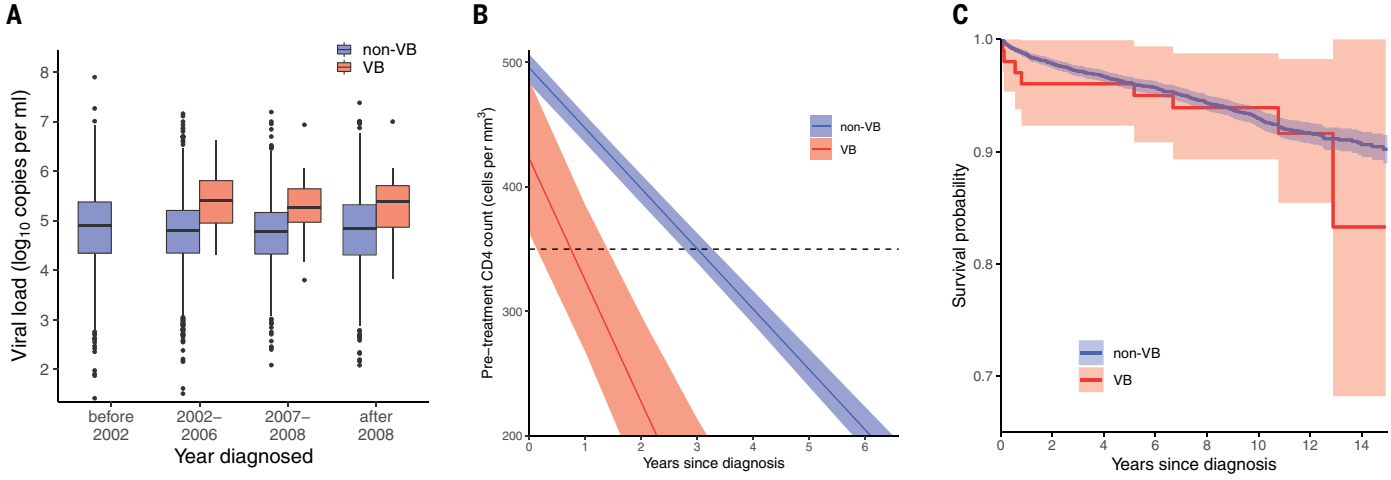


Fig. 1. Clinical characteristics of VB individuals. Those infected with the highly virulent variant (VB individuals) are represented in red; those infected with any other subtype-B virus (non-VB individuals) are shown in blue. (**A**) Box-and-whisker plots of viral load, by year of diagnosis. Diagnosis dates were grouped to produce boundaries that coincide with years and roughly equal numbers of VB individuals (39 in 2002–2006, 35 in 2007–2008, and 27 after 2008; the pattern is robust to other groupings). (**B**) Expected decline in CD4 count in the absence of treatment. The model was adjusted for sex and age at diagnosis; values shown are for males diagnosed at the age of 30 to 39 years. Shaded regions indicate 95% CIs in the model's prediction of mean values, given the uncertainty in estimation of parameter values (it does not reflect the variability between individuals in each of the two groups, which is much greater). The dashed black line denotes a CD4 count of 350 cells/mm^3 (see text for details). (**C**) Probability of still being alive at a given time after diagnosis.

(*34*). To provide context for these numbers, within Dutch subtype-B data in BEEHIVE we defined 16 other clades that are similar to the lineage in size (see materials and methods). For each clade, we calculated the amino acid consensus sequence, compared this to the Dutch subtype-B overall consensus, and determined CTL escape mutations. This showed that the number of such mutations for the VB variant is typical when normalized by its overall level of divergence (fig. S11). We also calculated the ratio of rates of nonsynonymous and synonymous changes (d_n/d_s) for each gene, for the VB variant and the other 16 Dutch subtype-B clades used for comparison. The VB variant had lower d_n/d_s values than all of the other clades for *env*, *pol*, and *tat*, though its values were not extreme; for the other genes, its d_n/d_s value was in the range spanned by the other clades (fig. S12). Finally, at codon position 77 of the protein Vpr, the consensus of all Dutch subtype-B sequences in BEEHIVE is glutamine, whereas the VB consensus is arginine. Glutamine was previously found to be more common in long-term nonprogressors, and mutation to arginine increased T cell apoptosis in vitro and strongly increased T cell decline in mouse models (*35*). However, both alleles have been commonly observed in subtype B to date (of 2178 subtype-B Vpr protein sequences in the Los Alamos National Laboratory HIV Database, 52% have glutamine and 36% have arginine), making it implausible that this mutation alone is the dominant mechanism for the virulence effect we observed.

Evolution of the VB variant

The maximum-likelihood phylogeny in Fig. 2A shows the VB variant in the context of background sequences, demonstrating that it is a distinct genetic cluster characterized by high viral loads. The phylogeny was inferred from 15 whole-genome VB-variant sequences and 100 randomly chosen whole-genome subtype-B background sequences from BEEHIVE. Figure 2B shows a dated phylogeny for VB-variant sequences only, estimated by using *BEAST* (*36*) and partial *pol* sequences. This phylogeny is colored by region, inferred with an ancestral state reconstruction by parsimony (minimizing changes of region). Amsterdam was assigned to the most recent common ancestor in 97% of trees in the posterior, showing that this reconstruction was robust to the uncertainty in the phylogeny. All VB-variant sequences date from 2003 onward; the time of their most recent common ancestor (TMRCA) was estimated as 1998.0 (95% credibility interval: 1995.7 to 2000.1). Trees were visualized by using *ggtree* (*37*).

Phylodynamics of the VB variant

The effective population size (N_e) of a pathogen is indicative of the number of infectious people. For the VB variant, N_e was estimated by using a sky grid demographic model (*38*) in *BEAST* and is shown in Fig. 2C (scaled by the coalescent generation time τ). N_e increased until roughly 2010; after this, there is more uncertainty but a possible downward trend [which may be an artefact of N_e inference methods in the recent past (*39*)]. The proportion of VB-variant cases among all new subtype-B cases increased until a peak in 2008 and subsequently decreased, though again with appreciable uncertainty [absolute numbers of both VB and non-VB diagnoses in our dataset have been decreasing since roughly 2008, and the data are right-censored by several years (fig. S7)]. In a recent analysis of an updated version of the ATHENA dataset (*40*), 33 additional VB individuals were found, which suggests that VB diagnoses were stable until roughly 2013 and have since been declining, still with appreciable uncertainty.

We calculated the local branching index (LBI), which is a measure of fitness (*41*). For HIV in a context in which most individuals start treatment without long delays, the LBI is closely related to transmissibility (see supplementary text). Compared with that of other transmission clusters, the LBI was higher for the VB variant both in BEEHIVE ($P = 2 \times 10^{-7}$) and ATHENA ($P < 2 \times 10^{-16}$; fig. S8). High pretreatment transmissibility may explain why the VB variant grew to be the 10th largest of 1783 clusters in the full ATHENA tree.

Tree imbalance and evolution within the VB-variant clade

We found nothing unusual in the extent to which the VB variant's phylogeny is imbalanced, nor did we detect any indication of further evolution of viral load within the variant's clade (supplementary text and fig. S9).

The first sampled VB individual

We retrieved and sequenced two additional samples from the VB individual who was diagnosed in 1992, 10 years before subsequent diagnoses

of other VB individuals. Phylogenetic analysis suggested that this individual was infected with a virus that had evolved most of the way, but not entirely, toward VB-variant viruses typical of later dates (supplementary text and fig. S10). This individual was diagnosed in Amsterdam, consistent with the aforementioned ancestral reconstruction of region. In the 10 years before this first VB diagnosis, the proportion of individuals diagnosed in the Netherlands for whom a viral sequence was available was roughly one-third. The proportion of those diagnosed or undiagnosed would be smaller still. This means that the infector of the 1992 individual was most likely not sampled, and indeed two or three steps in the transmission chain could have been unsampled. The long phylogenetic branch leading to the 1992 individual could therefore represent between-host evolution, not necessarily within-host evolution in a single individual.

Discussion

Previous studies of the heritability of viral load and CD4 cell decline led us to expect that these properties could change with the emergence of a new variant of HIV-1. We provided strong evidence for this, discovering a virulent subtype-B variant (the VB variant) that has been circulating in the Netherlands since the late 1990s. We characterized the variant's genotype and evolutionary history, as well as its association with high viral loads, rapid decline of CD4 cells, and increased transmissibility. We found 109 individuals with the variant (VB individuals) whose age, sex, suspected mode of transmission, and region of birth are all typical for people living with HIV in the Netherlands. This suggests that the observed association is causal: The increased virulence is a property of the virus rather than a confounding property of individuals in this transmission cluster. An absence of viral load evolution inside the clade of VB variants suggests that the increased virulence is a property of the whole clade and not a subset of it—i.e., that the virulence evolution occurred on the long phylogenetic branch that connects this clade to other known viruses.

Deferring the initiation of treatment until the measurement of a CD4 count's decline to ≤350 cells/mm^3 or the onset of AIDS, instead of immediate treatment initiation upon diagnosis, was previously shown to increase the subsequent hazard of serious AIDS-related events by a factor of 3.6 (CI: 2.0 to 6.7) and of any serious event (including death) by a factor of 2.4 (CI: 1.6 to 3.3) (*25*). This long-lasting immunological damage justifies WHO's classification of 350 CD4 cells/mm^3 as "advanced HIV." Without treatment, advanced HIV is expected to be reached in only 9 months (CI: 2 to 17) from the time of diagnosis for VB individuals, compared with 36 months (CI: 33 to 39) for non-VB individuals, in males diagnosed at the age of 30 to 39 years. Advanced HIV is reached even more quickly in older age groups, and there is considerable variation between individuals around these expected values. Many individuals could therefore progress to advanced HIV by the time they are diagnosed, with a poorer prognosis expected thereafter in spite of treatment. In practice, there is still substantial variation in the delay from becoming infected to starting treatment, making the VB variant a concern even in the high-awareness and highly monitored context of the Dutch HIV-1 epidemic. In contexts with less awareness and monitoring, in which diagnosis often occurs later in infection, the probability of reaching advanced HIV before diagnosis would be even greater.

Future in vitro investigations could more firmly establish the role of the viral genotype, and reveal an as-yet-unknown virulence mechanism at the molecular or cellular level. A higher replicative capacity of the virus might be observed, given the increased viral loads seen here. However, it is likely that there will be more to the virulence mechanism: The VB variant doubles the rate of CD4 cell decline, measured with both counts and T cell percentages, even after adjusting for its higher viral load. This rate is equivalent to the acceleration of CD4 degradation that

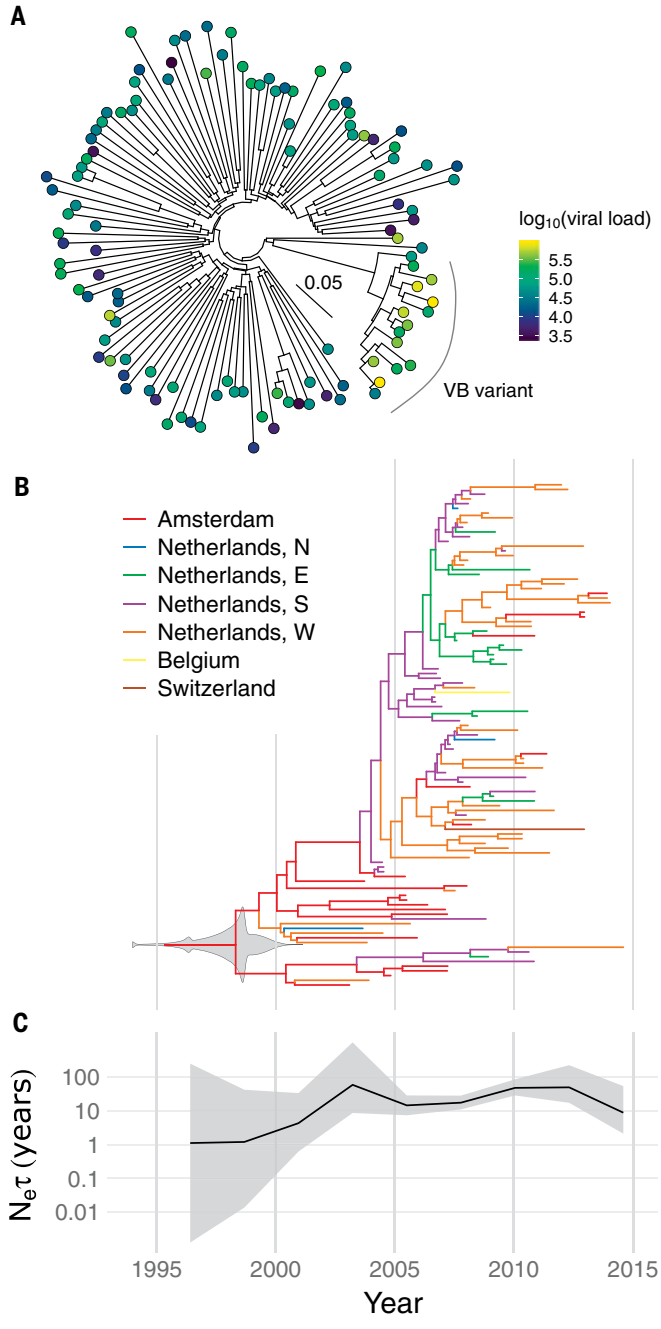


Fig. 2. Phylogenetic and phylodynamic analysis of the VB variant. (**A**) Whole-genome maximum-likelihood phylogeny of 15 VB-variant sequences and 100 background subtype-B sequences. The color of each circle indicates the individual's viral load in $\log_{10}$ copies per milliliter. The inset scale bar shows the branch length scale in units of substitutions per site. (**B**) Dated maximum-clade-credibility tree for 107 partial *pol* gene sequences from the VB variant. Colors indicate geographical regions (N, E, S, and W: north, east, south, and west), which are known for the tips of the tree but are otherwise inferred by ancestral state reconstruction. The gray violin plot superimposed on the root node shows the posterior density for its date (i.e., the TMRCA); 1994 contains overflow to earlier dates for clarity. (**C**) Effective population size (N_e) (scaled by the coalescent generation time τ) over time with 95% credibility intervals, with the same time axis as in (B).

would be expected from a 3.0 $\log_{10}$ increase in viral load, though we observed a 0.54 to 0.74 $\log_{10}$ increase. This means that the virulence normalized by the amount of virus—the "per-parasite pathogenicity" (*42*, *43*), which for HIV is heritable (*19*)—is much higher for the VB variant. Using two aforementioned methods, we predicted that, of the 17 whole genomes available, 16 use only the R5 co-receptor for cell entry, which is typical for subtype-B viruses in early infection (*13*). This finding suggests that the underlying virulence mechanism is distinct from the well-known effect of cell tropism (*14*, *15*).

Previous studies have reported population-wide increases (*44*, *45*) and decreases (*46*) in virulence over time. Mixed results between individual studies [see (*47*) for a meta-analysis] can be attributed to differences in epidemic context (such as the dominant subtypes), statistical power, and observational biases over time. Temporal virulence trends could also be due to changing confounders, such as a shift in which subpopulations are most affected, the stage of infection at time of diagnosis, or coinfections. We expand on these studies by resolving a change in virulence to an individual viral variant.

The basic theory of an infectiousness–virulence trade-off is that infectiousness and virulence are linked (for example, by how fast a pathogen replicates in its host) and that selection pressures favor intermediate values rather than extreme ones. If infectiousness is too low, the pathogen cannot be transmitted when its host contacts other hosts, but if virulence is too high, the host becomes too ill to have such contacts. In the case of HIV, the implication of this theory is that we would not expect highly virulent viruses to spread widely through a population in the absence of widespread treatment, because their hosts would progress to AIDS very quickly, limiting the opportunities for transmission (*9*). Most of the evolution that gave rise to the VB variant occurred before 1992, before effective combination treatment was available. However, our findings may stimulate further interest in whether widespread treatment shifts the balance of the infectious–virulence trade-off toward higher virulence, thus promoting the emergence and spread of new virulent variants. Previous modeling studies have investigated this idea for pathogens generally (*48*) and for HIV specifically (*49*, *50*). We discuss subtleties of the argument in the supplementary text, but our conclusion is that widespread treatment is helpful to prevent new virulent variants, not harmful. The absolute fitness of viral variants must be considered in addition to their relative fitness, and treatment reduces the total onward transmission over the course of one infection, regardless of virulence. Put simply, "viruses cannot mutate if they cannot replicate" (anonymous), and "the best way to stop it changing is to stop it" [Marc Lipsitch (*51*)]. Early treatment also prevents CD4 cell decline from leading to later morbidity and mortality; thus clinical, epidemiological, and evolutionary considerations are aligned. Our discovery of a highly virulent and transmissible viral variant therefore emphasizes the importance of access to frequent testing for at-risk individuals and of adherence to recommendations for immediate treatment initiation for every person living with HIV (https://www.who.int/health-topics/hiv-aids).

REFERENCES AND NOTES

1. J. L. Geoghegan, E. C. Holmes, *Nat. Rev. Genet.* **19**, 756–769 (2018).
2. C. E. Cressler, D. V. McLeod, C. Rozins, J. van den Hoogen, T. Day, *Parasitology* **143**, 915–930 (2016).
3. J. J. Bull, A. S. Lauring, *PLOS Pathog.* **10**, e1004387 (2014).
4. R. Challen *et al.*, *BMJ* **372**, n579 (2021).
5. N. G. Davies *et al.*, *Nature* **593**, 270–274 (2021).
6. D. J. Grint *et al.*, *Euro Surveill.* **26**, (2021).
7. N. G. Davies *et al.*, *Science* **372**, eabg3055 (2021).
8. E. Volz *et al.*, *Nature* **593**, 266–269 (2021).
9. C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, W. P. Hanage, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17441–17446 (2007).
10. P. M. Sharp, B. H. Hahn, *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011).
11. M. Worobey *et al.*, *Nature* **455**, 661–664 (2008).
12. J. Hemelaar *et al.*, *Lancet Infect. Dis.* **19**, 143–155 (2019).
13. B. S. Taylor, M. E. Sobieszczyk, F. E. McCutchan, S. M. Hammer, *N.Engl. J. Med.* **358**, 1590–1602 (2008).
14. B. Asjö *et al.*, *Lancet* **2**, 660–662 (1986).
15. M. Koot *et al.*, *Ann. Intern. Med.* **118**, 681–688 (1993).
16. H. Song *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 239–244 (2019).
17. C. Fraser *et al.*, *Science* **343**, 1243727 (2014).
18. F. Blanquart *et al.*, *PLOS Biol.* **15**, e2001855 (2017).
19. F. Bertels *et al.*, *Mol. Biol. Evol.* **35**, 27–37 (2018).
20. C. Fraser, T. D. Hollingsworth, *AIDS* **24**, 2596–2597 (2010).
21. V. Mitov, T. Stadler, *Mol. Biol. Evol.* **35**, 756–772 (2018).
22. F. Schoeni-Affolter *et al.*, *Int. J. Epidemiol.* **39**, 1179–1189 (2010).
23. K. Kusejko *et al.*, *J. Infect. Dis.* **220**, 244–253 (2019).
24. A. U. Scherrer *et al.*, *Int. J. Epidemiol.* 10.1093/ije/dyab141 (2021).
25. J. D. Lundgren *et al.*, *N. Engl. J. Med.* **373**, 795–807 (2015).
26. S. Lodi *et al.*, *Clin. Infect. Dis.* **53**, 817–825 (2011).
27. A. van Sighem *et al.*, *Epidemiology* **26**, 653–660 (2015).
28. A.-C. Pineda-Peña *et al.*, *Infect. Genet. Evol.* **19**, 337–348 (2013).
29. D. Struck, G. Lawyer, A.-M. Ternes, J.-C. Schmit, D. P. Bercoff, *Nucleic Acids Res.* **42**, e144 (2014).
30. M. A. Jensen *et al.*, *J. Virol.* **77**, 13376–13388 (2003).
31. M. C. Prosperi *et al.*, *AIDS Res. Hum. Retroviruses* **25**, 305–314 (2009).
32. P. Kellam, C. A. Boucher, B. A. Larder, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1934–1938 (1992).
33. S.-Y. Rhee *et al.*, *Nucleic Acids Res.* **31**, 298–303 (2003).
34. J. M. Carlson *et al.*, *J. Virol.* **86**, 13202–13216 (2012).
35. J. J. Lum *et al.*, *J. Clin. Invest.* **111**, 1547–1554 (2003).
36. M. A. Suchard *et al.*, *Virus Evol.* **4**, vey016 (2018).
37. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, *Methods Ecol. Evol.* **8**, 28–36 (2017).
38. M. S. Gill *et al.*, *Mol. Biol. Evol.* **30**, 713–724 (2013).
39. E. de Silva, N. M. Ferguson, C. Fraser, *J. R. Soc. Interface* **9**, 1797–1808 (2012).
40. D. Bezemer *et al.*, AIDS **36**, 83–94 (2022).
41. R. A. Neher, C. A. Russell, B. I. Shraiman, *eLife* **3**, e03568 (2014).
42. L. Råberg, M. Stjernman, in *Ecoimmunology*, G. Demas, R. Nelson, Eds. (Oxford Univ. Press, 2011), pp. 548–578.
43. L. Råberg, *PLOS Biol.* **12**, e1001989 (2014).
44. N. Pantazis *et al.*, *Lancet HIV* **1**, e119–e126 (2014).
45. J. O. Wertheim *et al.*, *Nat. Commun.* **10**, 5788 (2019).
46. F. Blanquart *et al.*, *eLife* **5**, e20492 (2016).
47. J. T. Herbeck *et al.*, *AIDS* **26**, 193–205 (2012).
48. T. C. Porco, J. O. Lloyd-Smith, K. L. Gross, A. P. Galvani, *J. Theor. Biol.* **233**, 91–102 (2005).
49. J. T. Herbeck *et al.*, *Virus Evol.* **2**, vew028 (2016).
50. H. E. Roberts, P. J. R. Goulder, A. R. McLean, *J. R. Soc. Interface* **12**, 20150888 (2015).
51. Marc Lipsitch (@mlipsitch), Twitter, 15 February 2020, https://twitter.com/mlipsitch/status/1228734360716791815.
52. C. Wymant, C. Fraser, F. Blanquart, Analysis code for "A highly virulent variant of HIV-1 circulating in the Netherlands", Zenodo, version 1 (2021); https://doi.org/10.5281/zenodo.5761935.

ACKNOWLEDGMENTS

We thank K. Fransen and G. Vanham for help with the Belgian data, O. Ratmann for help in identifying the Dutch clusters, K. Kusejko for testing for additional VB individuals in the SHCS, B. Foley for help with genome sharing, B. Dearlove and L. Thomson for help with software, and J. Herbeck and three other reviewers for helpful suggestions. **Funding:** This study was funded by ERC Advanced Grant PBDR-339251 and a Li Ka Shing Foundation grant, both awarded to C.F. The ATHENA Cohort is managed by Stichting HIV Monitoring and supported by a grant from the Dutch Ministry of Health, Welfare and Sport through the Centre for Infectious Disease Control of the National Institute for Public Health and the Environment. **Author contributions:** Funding acquisition: C.F. Conceptualization: C.W. and C.F. Data generation and management: all authors. Investigation: C.W., F.B., D.B., L.F., P.R., and C.F. Writing, original draft: C.W. and C.F. Writing, review and editing: all authors. **Competing interests:** P.K. is an employee of Kymab, a Sanofi company. H.F.G. reports grants from the Swiss National Science Foundation, National Institutes of Health (NIH), and the Swiss HIV Cohort Study; unrestricted research grants from Gilead Sciences, Roche, and the Yvonne Jacob Foundation; and personal fees from consulting or advisory boards or data safety monitoring boards for Merck, Gilead Sciences, ViiV Healthcare, Mepha, and Sandoz. H.F.G.'s institution received money for participation in the following clinical COVID-19 studies: 540-7773/ 5774 (Gilead), TICO (ACTIV-3, INSIGHT/NIH), and the Morningsky study (Roche). **Data and materials availability:** Code illustrating the analysis of the source clinical data, and of the genomic distribution and annotation of VB-variant mutations, is openly available at GitHub (https://github.com/ChrisHIV/hiv_vb_variant); a version has also been deposited at Zenodo (*52*). The 17 VB-variant whole genomes are publicly available at GenBank with accession numbers MT458931 to MT458935 and MW689459 to MW689470; the two putative recombinants have accession numbers MW689465 and MW689466. Data on viral loads, pretreatment CD4 counts, and mortality are provided as data S3. Requests for further data access can be made by submission of a concept sheet to the corresponding authors; these will be reviewed on a case-by-case basis, given that the data underlying this study contain sensitive and potentially identifying information. Once submitted, the proposed research and/or analysis will undergo review by the BEEHIVE Data Access Committee, which includes representatives of the ATHENA Cohort, for evaluation of scientific value, relevance to the study, design and feasibility, statistical power, and overlap with existing projects. If the proposed analysis is for verification and/or replication, data will then be made available. If the proposed research is for novel science, upon completion of the review, feedback will be provided to the proposer(s). In some circumstances, a revision of the concept may be requested. If the concept is approved for implementation, a writing group will be established that will consist of the proposers (up to three persons that were centrally involved in the development of the concept) and members of the BEEHIVE Collaboration and ATHENA Cohort (or other appointed cohort representatives). All persons involved in the process of reviewing these research concepts are bound by confidentiality. **Ethics statement:** At initiation, the ATHENA Cohort was approved by the institutional review board of all participating institutions. People beginning HIV care receive written material about participation in the ATHENA study and are informed by their treating physician of the purpose of data collection, after which they can consent verbally or opt out. Data are pseudonymized before being provided to investigators and may be used for scientific purposes. A designated quality management coordinator safeguards compliance with the European General Data Protection Regulation. Additional written informed consent was obtained for ATHENA participants enrolled in BEEHIVE for whole-genome sequencing.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk1688
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S7
References (53–85)
MDAR Reproducibility Checklist
Data S1 to S4

25 June 2021; accepted 4 January 2022
10.1126/science.abk1688

Section Two: Original Content

New content provided by Abbott and its Pandemic Defense Coalition partners

Pandemic defense: The key is global collaboration

Gavin Cloherty*

As the world adapts to its third year of living with COVID-19, one thing has become clearer than ever: Viruses are in a race with science, and scientists must find ways not just to keep pace but to move faster. We have now entered an era where epidemic and pandemic threats have become the new normal.

Scientists need creative and innovative ways to hunt, track, and discover future viral threats as well as those facing us today. From highly transmissible COVID variants to deadly flu strains and more, viruses move fast and evolve even faster. *And so must we.*

It's a tall order, but success can be achieved through a combination of science and collaboration.

This means that we must take the lessons we've learned from the current pandemic and translate them into action, which includes developing strong partnerships between government and private companies; creating lasting coalitions of doctors, scientists, and researchers to watch and respond to outbreaks as they occur; and helping to train the next generation of epidemiologists.

Creating these partnerships between the private sector, academia, and governments can ensure we are ready for the next disease that comes our way. With 10 million times more viruses on Earth than stars in the universe,[1] there's no shortage of pathogens that could cause the next public health crisis.

During the COVID-19 pandemic, private companies stepped up to rapidly develop and manufacture tests. At the time of the original Omicron variant surge, there were nine at-home rapid COVID-19 tests approved by the U.S. Food and Drug Administration—and 375 million tests were available in the market[2]—but the volume was still not enough to meet household needs. It's critical that governments continue to work with private companies to procure key materials, secure manufacturing capacity, and coordinate a public health response strategy to raise pandemic defenses.

To be as efficient as possible, the public and private sectors will have to collaborate across continents and oceans, too. Viruses know no borders, so neither should our efforts to combat them. That's why Abbott, a global health care company building on a nearly 30-year history in viral surveillance, formed the Abbott Pandemic Defense Coalition—a first-of-its-kind, industry-led global scientific, academic, and public health partnership dedicated to improving early detection and rapid response to viral threats. The coalition, comprising 14 scientific, academic, and public health organizations located around the world, is actively working to help prevent the next pandemic. It works with partners on the ground to track SARS-CoV-2, hepatitis, HIV, and other global viruses by analyzing millions of viral sequences and screening samples

*Head of Infectious Disease Research and the Pandemic Defense Coalition, Abbott, Abbott Park, IL

isolated from patients with countless unknown illnesses in order to stay one step ahead of the next threat, so we can respond quickly and prevent widespread impact.

If a potential threat is discovered, the sequence of the viruses is published in a public database so that health officials and laboratories can work together to identify if it's a novel strain, or a virus that has been previously detected.

That approach proved effective when the COVID-19 Omicron variant popped up in South Africa late last year and again with new subvariants arising across the globe this year. Our partners in Africa shared variant sequences and tested samples, allowing Abbott to quickly confirm that our COVID-19 tests could detect it.

When faced with a new viral threat, Abbott will develop diagnostic testing to assist in containment efforts. The goal is to anticipate viral threats before they occur, enabling rapid responses and diagnostic testing to track how these threats are emerging and prevent them from becoming pandemics.

To fight the current pandemic from all sides, Abbott developed and produced 14 types of COVID-19 tests for use in hospitals, laboratories, point-of-care facilities, and homes. Further, because we knew the virus would mutate, we immediately included SARS-CoV-2 in our surveillance program so that we could study, sequence, and understand how it was changing.

Today, coalition partners are putting this model to the test by actively collecting, logging, and monitoring diverse pathogens to keep pace with new and changing viruses in every region of the globe—from rainforests in Colombia, Brazil, and Thailand, to the expansive savannas of South Africa, to the metropolitan areas of India and the United States.

In just the past year, our coalition has identified numerous new lineages of the SARS-C0V-2 virus. Gaining insight into the mutation rate over time, we've ensured that our molecular, antigen, and serological tests can detect variants; studied immune systems after vaccination to better understand COVID-19 recovery; and discovered other undetected pathogens in circulation in member countries, alerting public health officials so that proper testing and monitoring can be put in place.

As unknown viral threats become more common, we'll need more eyes on the ground. That's why we and the coalition are committed to growing and supporting the ranks of global virus hunters. Abbott is helping increase the number of future virus hunters through the coalition's partnership with the Training Programs in Epidemiology and Public Health Interventions Network (TEPHINET), which is already training field epidemiologists in more than 100 countries. And we'll continue supporting scientists at each of our coalition sites to develop infrastructure and skills for future response.

By expanding training programs, we can attract more virus hunters to be educated at research sites across the world and learn to respond more effectively in the fight against infectious diseases. A part of Abbott's sustainability plan is to build capacity and scientific skills today—which will enable us to strengthen pandemic preparedness and protect communities against the threats of tomorrow.

In an increasingly connected and interdependent world, it is imperative to streamline pandemic detection and response, improving links between international networks, governments, and public health organizations that are involved in global health decisions.

As we make our way through yet another year of living with COVID-19, we have to keep taking steps to stop the next pandemic. The time to work together to stop future viral threats is now. And collaboration is key.

[1]https://www.smithsonianmag.com/air-space-magazine/there-are-more-viruses-earth-there-are-stars-universe-180974433.

[2]https://www.whitehouse.gov/briefing-room/statements-releases/2022/01/14/fact-sheet-the-biden-administration-to-begin-distributing-at-home-rapid-covid-19-tests-to-americans-for-free.

Abbott

HELPING TO PREVENT THE NEXT PANDEMIC: THE ABBOTT PANDEMIC DEFENSE COALITION

HOW DO WE STOP HISTORY FROM REPEATING ITSELF?

This first-of-its-kind* global network is dedicated to detecting and identifying viral threats. Meet the teams who are working around the world to help prevent the next pandemic.

Abbott
Rush University
CHICAGO, U.S.

Colombia/Wisconsin One-Health Consortium at the Universidad Nacional de Colombia
MEDELLIN, COLOMBIA

National Center for Disease Control and Public Health
TBILISI, GEORGIA

YRG Care
CHENNAI, INDIA

Fiocruz
RIO DE JANEIRO, BRAZIL

Centre for Epidemic Response and Innovation
STELLENBOSCH, SOUTH AFRICA

KRISP, Genomic Centre of the University of KwaZulu-Natal
DURBAN, SOUTH AFRICA

Faculty of Medicine, Siriraj Hospital, Mahidol University
BANGKOK, THAILAND

Universidade de São Paulo
SAO PAULO, BRAZIL

Université Quisqueya
PORT-AU-PRINCE, HAITI

Uganda Virus Research Institute
ENTEBBE, UGANDA

Institut de Recherche en Santé, de Surveillance Epidémiologique et de Formations (IRESSEF)
DAKAR, SENEGAL

The University of the West Indies
MONA, JAMAICA

University of Zimbabwe
HARARE, ZIMBABWE

HOW IT WORKS: THE COALITION'S ACTION PLAN

1 When doctors identify patients with unknown conditions that they can't diagnose, it can be the first sign of a new threat.

2 The patient samples are sent to a coalition member, where they'll undergo further analysis to determine if we've seen the virus before.

3 We'll conduct genetic sequencing, which will identify characteristics that may indicate a new threat.

4 If a threat is discovered, the viral sequence and prototype diagnostic test will be immediately shared so that science and public health can collaborate to understand the risk level and pandemic potential.

5 We'll deploy diagnostic testing to the extent possible to assist in efforts to contain the virus before it can reach pandemic levels.

*The Abbott Pandemic Defense Coalition is the first company-led global scientific and public health network dedicated to helping prevent and diminish future pandemic threats. www.abbott.com/virushunters.html

OVERVIEW

The Abbott Pandemic Defense Coalition connects organizations around the world that contribute to global health, scientific research and pandemic preparedness. Whether they're conducting viral surveillance and discovery, analyzing genomic sequences, or training the next generation of epidemiologists and virologists, all partners play a vital role in staying a step ahead of infectious disease outbreaks.

AMERICAS

Rush University Medical Center

Rush is one of the leading academic health systems in the U.S. and works to discover novel treatments through research and clinical trials.

Specialization: Clinical Research, Immunology, Informatics, Next-Generation Sequencing, Training and Education

Universidade de São Paulo (USP)

USP is a leading university and plays a fundamental role in advancing scientific, technological and social research in Brazil.

Specialization: Clinical Research, Health Geography, Next-Generation Sequencing, Phylogenetics

Colombia-Wisconsin One Health Consortium at the Universidad Nacional de Colombia

University of Wisconsin–Madison and Universidad Nacional de Colombia strengthen scientific capabilities with one of the top laboratories outfitted for genomic technology in the region.

Specialization: Clinical and Epidemiological Research, Diagnosis, Health Geography, Next-Generation Sequencing

University of the West Indies

The University of the West Indies aims to advance learning, create knowledge and foster innovation for the positive transformation of the Caribbean and the world.

Specialization: Genome Sequencing, Health Geography, Next-Generation Sequencing, Vector Surveillance, Viral Surveillance

Fiocruz

Fiocruz, or the Oswaldo Cruz Foundation under the Ministry of Health, is the most prominent science and technology institution in health in Latin America.

Specialization: Clinical Research, Epidemiological Surveillance, Next-Generation Sequencing, Phylogenetics, Training and Education, Virology

Université Quisqueya

Université Quisqueya is a private education and research institution that is the first of its kind in Haiti.

Specialization: Clinical Research, Training and Education, Viral Surveillance

AFRICA

Uganda Virus Research Institute (UVRI)

UVRI conducts health research pertaining to human infections and diseases and enables partnerships and communication while serving as a training and education center.

Specialization: Entomology, Epidemiological Surveillance, Immunology, Training and Education, Virological Research

University of Zimbabwe

This public research institution provides academic programs, research, and advisory services while developing strategic partnerships for innovative research, outreach and business development.

Specialization: Clinical Research, Phylogenetics, Training and Education, Viral Surveillance

Centre for Epidemic Response and Innovation (CERI)

CERI uses its expertise in genomics to identify new variants of pathogens and to advance basic and translational science to improve public health efforts in Africa.

Specialization: Bioinformatics, Genome Sequencing, Phylogenetics, Training and Education

Institut de Recherche en Santé, de Surveillance Epidémiologique et de Formation (IRESSEF)

An institution known for health research, IRESSEF aims to support public health policies to make healthcare accessible to African populations.

Specialization: BSL-3, Epidemiological Surveillance, Clinical Research, Virus Discovery

KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Genomic Centre of the University of KwaZulu-Natal

KRISP utilizes big data, bioinformatics and genomics analysis to understand the infectious diseases of today to better respond to future mutations or variants.

Specialization: Bioinformatics, Genome Sequencing, Phylogenetics, Training and Education

EUROPE AND ASIA

National Center for Disease Control and Public Health (NCDC)

The NCDC is Georgia's central public health and research institution under the authority of the Ministry of Internally Displaced Persons (IDPs) from the Occupied Territories, Labour, Health and Social Affairs.

Specialization: Clinical Research, Connecting Laboratories and Physicians, Countrywide Surveillance, Hepatitis Elimination Programs, Informatics

Faculty of Medicine, Siriraj Hospital, Mahidol University

Our partners at Siriraj Hospital, the leading medical school in the country, provide the latest medical services support and education while driving innovative infectious disease research.

Specialization: Bioinformatics, Biorepository, Clinical Research, Diagnosis, Health Geography

Y.R. Gaitonde Centre for AIDS Research and Education (YRG Care)

YRG Care carries an international reputation for clinical, behavioral and epidemiological research with global partners and provides community services for people living with HIV, AIDS and STIs.

Specialization: Clinical Research, Epidemiology, Patient Care Services for High-Risk Populations, Viral Surveillance

Protecting public health through viral molecular surveillance

Mary A. Rodgers*

The continuous evolution and divergence of viral pathogens, which are both fascinating and deadly, is constantly threatening the advances we have made in combating infectious diseases through diagnostics, therapeutics, and vaccines. All these tools rely upon viral sequence conservation to maintain effectiveness, making them prone to failure when viruses mutate. The magnitude of the challenge of viral diversity is most notable in the case of highly divergent viruses such as human immunodeficiency virus (HIV), hepatitis B virus (HBV), and hepatitis C virus (HCV), which have diversified into lineages that differ by up to 37.5% and are responsible for 38 million, 257 million, and 71 million chronic infections worldwide, respectively (*1*–*4*). With increasing global travel, immigration, and military deployments, the geographical footprint of these viruses is expanding (*5*, *6*). Although SARS-CoV-2 has lower overall viral diversity than HIV, HBV, or HCV, the pandemic spread of variants of concern around the world in a matter of weeks has repeatedly altered local transmission dynamics, making small changes in sequence a major driver of COVID-19 illness (*7*). Therefore, the comprehensive spectrum of viral genetic diversity must be accounted for in the design and evaluation of diagnostics, therapeutics, and vaccines targeting these and other divergent viruses. The consequences of an incomplete consideration of viral diversity include false negative diagnostic results; underquantification of viral load levels; vaccine breakthrough infections; treatment failures; and ultimately poor control of infectious diseases (*8*–*16*).

Molecular surveillance serves as the foundation for the robust design of diagnostics, therapeutics, and vaccines to effectively confront viral diversity. Complete characterization of the full diversity of pandemic viruses such as HIV, HBV, HCV, and SARS-CoV-2 requires a global effort, since even one gap in the worldwide molecular surveillance system puts

*Principal Research Scientist, Abbott Pandemic Defense Coalition, Abbott, Abbott Park, IL

the global population at risk. In addition to sequences from surveillance programs, positive clinical specimens are also critically important for continuously monitoring the impact of viral diversity on diagnostic tests. These paramount needs were recognized over 28 years ago when the Abbott Global Surveillance Program (AGSP) was established in 1994, with the goal of characterizing HIV diversity to improve the performance of diagnostic tests. The program grew out of decades of leadership in infectious disease diagnostics that began with Abbott's first test for HBV in 1972 and the first FDA-approved test for HIV in 1985. AGSP has since expanded to collecting more than 90,000 clinical specimens from 46 different countries on six continents, through collaborations with hospitals, blood banks, research institutions, and public health agencies (*17*, *18*). Access to the data generated by the program has been democratized for the benefit of all researchers through deposition of more than 5,000 new viral sequences into public databases and the publication of 152 peer-reviewed articles. Furthermore, AGSP-derived sequences and specimens serve as the foundation for the development of infectious disease diagnostic assays that can tolerate viral diversity; these assays serve as our critical first line of defense in the efforts to contain and eliminate HIV, viral hepatitis, and SARS-CoV-2. As a key aspect of pandemic preparedness, AGSP serves as a pillar of the Abbott Pandemic Defense Coalition.

Over the course of decades of molecular surveillance, AGSP has made several important discoveries in HIV research. A majority of the most divergent HIV strains have been identified and sequenced by AGSP, including 11 out of 17 sequence-confirmed HIV Group N infections identified to date (*19–24*) and one of the only two Group P cases ever to have been discovered (*25*). Numerous HIV-2 and HIV-1 Group O strains have also been characterized by AGSP, some of which differ from Group M sequences in the envelope gene—a primary target of the antibody response to HIV—by >50% (*20, 26–29*). More recently, the newest subtype of HIV Group M, subtype L, was established with the third genome sequenced by AGSP in 2019 (*30*). These diverse HIV strains illustrate the impact of a robust viral surveillance program on diagnostic assay development, since Abbott's ARCHITECT HIV Ag/Ab Combo and RealTi*m*e HIV-1 tests were designed with awareness of viral diversity prior to the discoveries of these strains, yet they were all readily detected without modification of the tests (*21*, *30*).

AGSP has also brought new insights and understanding about the course of the HIV pandemic (*31*). AGSP's HIV surveillance in Africa has uncovered an unexpected population of potential HIV elite controllers in the Democratic Republic of Congo (DRC) who can control the virus naturally without suppression by antivirals (*32*). The discovery of this group of patients was an unexpected outcome of a viral surveillance study in which the frequency of samples with viral load was unusually low in a seropositive HIV panel from Kinshasa. Through systematic follow-up testing, AGSP eliminated diagnostic errors, divergent viruses, and antiretrovirals as potential explanations. While the frequency of HIV elite controllers is typically less than 0.5% of the population in other regions of the world, their prevalence in DRC was estimated to be 2.7%–4.4%, a figure that remained consistent over a 30-year period, suggesting that a unique capacity to control HIV infection arose early in the epicenter of the pandemic. These results serve as a spark for renewed investigation in this area to characterize the mechanisms by which potential elite controllers in DRC achieve viral suppression.

After the addition of hepatitis virus surveillance into AGSP in the 2000s, several important contributions have been made in molecular surveillance of HBV, HCV, and hepatitis D virus (HDV) that have benefited public health. In a large

study conducted with >13,000 participants in Cameroon, where the prevalence of viremic HCV was previously only modeled and not measured, the molecular prevalence of HCV was determined to be 2.47%, identifying higher rates in males (3.81%) and persons over the age of 40 (9.09%) (*33*). In this same group, a high prevalence of HBV infections was found (10.45%), and, among these, the prevalence of HDV antibodies was 46.73% (*20*, *34*). Remarkably, HDV RNA was found in 34.2% of all HBV^+ cases, with the highest rates found in children ages 2–17 years old (*34*).

The sequences generated through molecular surveillance studies can also provide deeper insight into the transmission dynamics of a countrywide epidemic. For example, in a 2020 AGSP molecular surveillance study conducted in a group of people who inject drugs (PWID) in five cities in India, 10 different subgenotypes were present, whereas significantly less diversity was seen for HIV in the same cohort, with only one major strain predominating (subtype C) (*35*). The distribution of the HCV genotypes across different cities suggested novel introductions of HIV and HCV into PWID populations consistent with known drug trafficking routes along nearby borders. Spatiotemporal phylodynamic analysis of the HCV sequences from this study found genetic links within transmission networks between geographically distant Indian cities, indicating that targeted interventions must account for intercity travel to be successful (*36*).

AGSP has kept pace with the emergence of SARS-CoV-2 variants during the COVID-19 pandemic by characterizing specimens from four continents and testing them in 11 different antigen, molecular, and antibody COVID-19 assays, with excellent performance demonstrated due to well-conserved assay targets and robust design (*37*). All major variants of concern to date (Alpha, Beta, Gamma, Delta, and Omicron) have been sequenced in the program, with genomes deposited in the Global Initiative on Sharing All Influenza Data (GISAID) repository (*38*). Important insights about the pandemic have also been gained from AGSP-derived SARS-CoV-2 genomes. Notably, genomes carrying spike protein receptor-binding domain (RBD) mutations linked to increase transmission and vaccine escape in variants of concern were found in Senegal during the second wave of the pandemic in the winter of 2020–2021 (*39*). The mutations were found in lineages that were not designated as—and never became—variants of concern, raising questions about the path a variant must take to become a global concern. Interestingly, the Alpha variant of concern was present during the second wave, but it never became the predominant strain in Senegal in contrast to other countries where it rapidly became the primary lineage. This may be due to the pre-existing dominance of the B.1.1.420 lineage in Senegal at the time that Alpha was introduced (*40*). The B.1.1.420 lineage was imported from Italy and gained fitness in Senegal before spreading globally during the second wave, making it a lineage to continue monitoring carefully (*40*).

The continual discovery of new strains and deeper insights gained from viral sequences in AGSP highlight the importance of sustained vigilance in monitoring viral sequences through a global approach to keep pace with viral evolution. To further improve our coverage of circulating strains, techniques that increase sequencing throughput and quality should continue to be implemented in as many geographies as possible. For instance, advances in improving next-generation sequencing (NGS) depth and coverage for low viral load specimens have been developed at Abbott through

enrichment strategies for SARS-CoV-2, HIV, and hepatitis viruses. Implementation of these techniques at additional sites will increase the number of genomes generated (*41–44*). Expanded application of surveillance methods to passively monitor the environment will also improve real-time visibility of strains that are circulating in specific areas. Notably, recent implementation of wastewater monitoring of SARS-CoV-2 sequences has been shown to successfully predict surges in cases and could be used to prompt public health responses when new variants are detected (*45*, *46*). However, passive surveillance systems do not exclusively monitor human pathogens, and environmental specimens cannot be used to monitor diagnostic assay performance. As more and more sequencing data is generated through surveillance efforts, efficient analysis of the data will continue to pose a challenge, and innovative solutions in this area will greatly improve surveillance throughput and accessibility. Resources that provide up-to-date status reports pulled from surveillance databases will also become an essential part of the suite of monitoring tools used to make sense of the data generated. The more that surveillance sequencing data is shared, the stronger our defenses will be against emerging strains of divergent viruses such as HIV, hepatitis, and SARS-CoV-2. Since new variants of any of these viruses could cause major outbreaks and surges globally, improved surveillance of these pathogens is a critical pillar of pandemic preparedness. As the only diagnostic test manufacturer with such a unique long-standing and large-scale surveillance program, Abbott provides a vital tool to stay ahead of these dynamic viruses.

References

1. *Global Hepatitis Report, 2017*. (World Health Organization, Geneva); available at https://www.who.int/publications/i/item/9789241565455.
2. A. Kramvis, *Intervirology* **57**, 141–150 (2014).
3. World Health Organization, HIV Factsheet. (2018).
4. G. Li *et al.*, *Retrovirology* **12**, 18 (2015).
5. A. Beloukas *et al.*, *Infect. Genet. Evol.* **46**, 180–189 (2016).
6. A. M. Oster *et al.*, *Ann. Epidemiol.* **27**, 244–251 e241 (2017).
7. F. Campbell *et al.*, *Euro. Surveill.* **26**, 2100509 (2021).
8. K. Korn *et al.*, *J. Clin. Micro.* **47**, 1238–1240 (2009).
9. S. Chevaliez, M. Bouvier-Alias, R. Brillet, J. M. Pawlotsky, *Hepatology* **46**, 22–31 (2007).
10. H. H. Kessler *et al.*, *J. Clin. Vir.* **58**, 522–527 (2013).
11. B. Hendrickson *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **67**, 311–312 (2018).
12. H. LaRue, L. Rigali, J. M. Balada-Llasat, P. Pancholi, *J. Clin. Micro.* **50**, 1769–1772 (2012).
13. S. M. Alavian, W. F. Carman, S. M. Jazayeri, *J. Clin. Vir.* **57**, 201–208 (2013).
14. X. Qiu *et al.*, *J. Clin. Vir.* **92**, 62–68 (2017).
15. S. Fourati *et al.*, *Hepatology* **69**, 513–523 (2019).
16. S. Iketani *et al.*, *Nature* **604**, 553–556 (2022).
17. C. A. Brennan *et al.*, *J. Med. Vir.* **78**, S24–S29 (2006).
18. F. Averhoff *et al.*, *Int. J. Infect. Dis.* **117**, 356–360 (2022).
19. P. Bodelle *et al.*, *AIDS Res. Hum. Retroviruses* **20**, 902–908 (2004).
20. M. A. Rodgers *et al.*, *Virology* **504**, 141–151 (2017).
21. M. A. Rodgers *et al. AIDS Res. Hum. Retroviruses* **34**, 314–318 (2018).
22. A. Vallari *et al.*, *AIDS Res. Hum. Retroviruses* **26**, 109–115 (2010).
23. J. Yamaguchi *et al.*, *AIDS Res. Hum. Retroviruses* **22**, 83–92 (2006).
24. J. Yamaguchi *et al.*, *AIDS Res. Hum. Retroviruses* **22**, 453–457 (2006).
25. A. Vallari *et al.*, *J. Virol.* **85**, 1403–1407 (2011).
26. L. G. Gürtler *et al.*, *J. Virol.* **68**, 1581–1585 (1994).
27. J. Yamaguchi *et al.*, *AIDS Res. Hum. Retroviruses* **20**, 944–957 (2004).
28. J. Yamaguchi *et al.*, *AIDS Res. Hum. Retroviruses* **24**, 86–91 (2008).
29. C. A. Brennan *et al.*, *J. Acquir. Immune Defic. Syndr.* **49**, 432–439 (2008).
30. J. Yamaguchi *et al.*, *J. Acquir. Immune Defic. Syndr.* **83**, 319–322 (2020).
31. D. M. Junqueira *et al.*, *AIDS Res. Hum. Retroviruses* **36**, 574–582 (2020).
32. M. G. Berg *et al.*, *EBioMedicine* **65**, 103258 (2021).
33. M. A. Rodgers *et al.*, *J. Vir. Hep.* **26**, 30–37 (2019).
34. E. K. Butler *et al.*, *Sci. Rep.* **8**, 11617 (2018).
35. M. A. Rodgers *et al.*, *Sci. Rep.* **10**, 7214 (2020).
36. S. J. Clipman *et al.*, *Hepatology* **74**, 1782–1794 (2021).
37. M. A. Rodgers *et al.*, *J. Clin. Vir.* **147**, 105080 (2022).
38. S. Elbe, G. Buckland-Merrett, *Glob. Chall.* **1**, 33–46 (2017).
39. A. D. Ahouidi *et al.*, *Sci. Rep.* **11**, 23644 (2021).
40. L. J. Perez *et al.*, *Virus Evol.*, **8**, veac025 (2022).
41. J. Yamaguchi *et al.*, *Front. Microbiol.* **9**, 2150 (2018).
42. G. S. Orf *et al.*, *Front. Virol.*, **1**, 765974 (2021).
43. M. G. Berg *et al.*, *PLOS One* **15**, e0236046 (2020).
44. K. Forberg *et al.*, *Virology* **539**, 69–79 (2020).
45. S. Agrawal *et al.*, *Water Res.* **214**, 118162 (2022).
46. A. E. Kirby *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **71**, 103–105 (2022).

Virus discovery for diagnostic preparedness

Michael G. Berg*

The Abbott Pandemic Defense Coalition (APDC) is a scientific and public health partnership whose primary purpose is the early detection and mitigation of infectious disease threats of pandemic potential. To this day, a large percentage of presumed infectious diseases go undiagnosed. For example, 15%–30% of acute respiratory illness, 30%–50% of gastroenteritis, and 50%–70% of meningitis/encephalitis have no etiology determined (*1*). In this context and in contrast to "surveillance" of known pathogens, "discovery" refers to detection of novel, emerging viruses that have not been previously characterized or associated with disease. Unbiased next-generation sequencing (NGS) of patient samples has the potential to dramatically improve these numbers, but there is spirited discussion about its actual diagnostic relevance and actionability in clinical settings (*2*–*4*).

*Associate Research Fellow, Abbott, Abbott Park, IL

However, the question as to whether NGS and associated bioinformatic algorithms [sequence-based ultrarapid pathogen identification (SURPI), Kraken, IDseq, among others] can be used to discover the presence of and assemble the genomes of new viruses is not a matter of debate. In recent decades, various new viruses have been discovered by NGS with clear links to disease (*1*, *5*). This number is dwarfed by those described in nonhuman host reservoirs (e.g., bats and rodents) and insects (e.g., ticks and mosquitos) with potential for zoonosis or vectoring, respectively (*6*–*12*). One need not touch a sample, with new viruses being found by combing through the vast reams of data in the sequence read archive (*13*). It is important to appreciate that beyond determining the sequence, numerous other criteria need to be established to link a pathogen with disease, including elicitation of a host response, clustering of cases, and fulfillment of Koch's postulates (*14*).

During an outbreak, finding other cases and making the correct association with disease ceases to be an issue. Yet, distributing the technical NGS capability to make that initial identification of a new virus and to monitor strain diversity following its emergence are areas that we can focus on for pandemic preparedness. Whether it's due to globalization, international travel, changes in the environment, or encroachment of humans into natural habitats, pandemic-scale viruses such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Zika, Ebola, and SARS-CoV-2 have increased in frequency, and it is likely that other outbreaks with pandemic potential will emerge again. Diagnostics play the most essential role throughout a public health response to a pandemic. Quickly designing molecular and serologic/rapid tests can be done by any entity, but only industry has the resources, technical knowledge, and logistic infrastructure to scale up production rapidly enough to meet the global demand. It all starts with having a genome sequence. For this reason, Abbott's Virus Discovery program performs a necessary function within the APDC framework to bolster public health efforts (*15*).

The evolution of virus discovery at Abbott

The Abbott Diagnostics division was established in 1972 with the development and distribution of the AUSRIA-125 blood screening test for the hepatitis B virus (HBV) Australian antigen. During this period, the search for the elusive non-A, non-B hepatitis virus was at its height. Around the same time that Chiron Corporation discovered hepatitis C virus (HCV) in 1989, Abbott scientists identified another group of flaviviruses that appeared to be associated with hepatitis and other transfusion-transmitted diseases. During the 1990s, GB virus type C (GBV-C) was ultimately found to be ubiquitous in both healthy and sick individuals, and so while a new *Pegivirus* species was established, these viruses were not pathogenic (*16*). The pace of discovery was hampered by reliance on labor-intensive methods available at the time, such as cloning and Sanger sequencing, radio-labeling, and virus propagation in culture. The advent of high-throughput methods such as microarrays for viruses (ViroChip) held great promise for the virus discovery field. In 2006, studies using the ViroChip indicated that xenotropic murine leukemia virus-related virus (XMRV) was associated with prostate cancer (PC) (*17*, *18*). A flurry of follow-up reports additionally suggested a causal role for XMRV in chronic fatigue syndrome (CFS) (*19*, *20*). Numerous independent studies failed to confirm an association of XMRV with PC and CFS, and Abbott scientists created molecular and serologic tools that ultimately helped to refute these claims (*21*, *22*). While XMRV was determined to be a real virus arising from recombination of endogenous murine proviruses during passage of a human PC tumor in immunocompromised mice, the field at large concluded that the initial results were due to laboratory contamination with infected cell lines (*23*, *24*). Important scientific lessons were learned from both "discoveries" and highlighted the many pitfalls and challenges associated with this work. Abbott not only remained engaged in this space but committed to taking the lead.

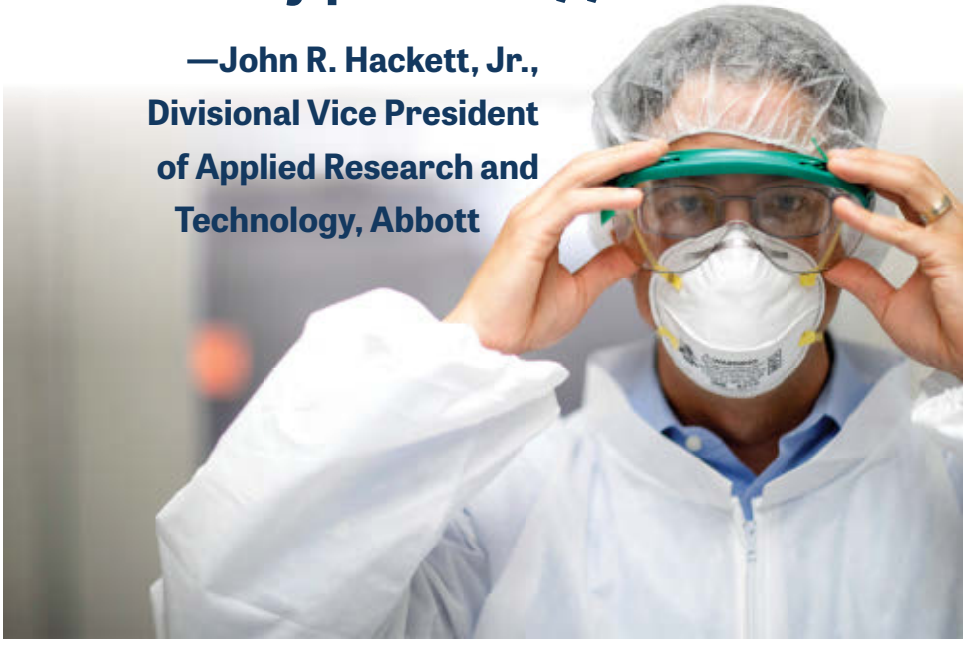
At this same time, the high-throughput sequencing revolution was underway with 454 and Illumina technologies quickly eclipsing microarrays. The concept of "metagenomic NGS" (mNGS) was born: unbiased sequencing of all nucleic acids in a patient sample to detect bacteria, parasites, fungi, and viruses, and to potentially assemble the genomes of new viruses. Abbott engaged in a unique, long-term partnership with the leading experts in this burgeoning field at University of California, San Francisco (UCSF) to establish the Viral Diagnostics and Discovery Center (VDDC). Leading the VDDC was Charles Chiu, M.D., Ph.D. Among his lab's many foundational contributions was leveraging mNGS to report the association of known viruses (e.g., astrovirus) with unexpected disease outcomes (e.g., encephalitis) (*25*). They also discovered several novel viruses, including MW polyomavirus in acute pediatric diarrhea cases, Bas-Congo rhabdovirus causing hemorrhagic fever, and Lone Star bunyavirus in *Amblyomma*

ticks (*26–28*). Importantly, this research demonstrated the promise of mNGS to deliver an actionable diagnosis, saving a young teenager from a fatal case of neuroleptospirosis (*29*). Pivotal to all this work was the development of the SURPI data-analysis pipeline (*30*). Vast amounts of sequences could now be aligned to the entire GenBank NT (Nucleotide) database in a few short hours, removing host background DNA/RNA, identifying known pathogens, and exploring clues as to the presence of a potential divergent virus. As intended, the focus then shifted to building NGS and bioinformatics capability internally at Abbott. A key milestone that evidenced this transfer of expertise and methodology was the discovery of human pegivirus-2 (HPgV-2) in 2015 (*31*). Starting from just three sequence reads with remote homology to simian pegivirus A, Chiu's group identified the index case and assembled the HPgV-2 draft genome. It was Abbott's team that developed this story further by identifying another dozen cases of infection, made possible by our ability to rapidly prototype molecular and serology tools. A quantitative reverse transcription (RT)-PCR on the Abbott *m*2000 RealTi*me* System simultaneously detecting HPgV-1 (GBV-C) and HPgV-2 viral RNA, together with a high-throughput assay on the ARCHITECT immunoassay analyzer detecting antibodies to envelope and NS4A/B antigens, yielded important epidemiologic data (*32*, *33*). We established that HPgV-2 was a rare, chronic bloodborne virus strongly associated with hepatitis C coinfection and likely transmitted by intravenous drug use. Concurrent studies from Kapoor *et al.* and other subsequent reports around the world corroborated this conclusion (*34–36*).

Stopping the next pandemic requires us to follow as many leads as possible. To do this effectively demands the building and optimization of methods and the development of the next generation of virus hunters. Recovering full genomes and comprehending viral genetic diversity from individuals

"Collaboration through VDDC laid an important foundation for Abbott's Pandemic Defense Coalition by acclimating us to the power of next-generation sequencing technology. This important partnership with UCSF allowed Abbott to build research capacity internally while also playing an active role in the virus discovery process."

—John R. Hackett, Jr., Divisional Vice President of Applied Research and Technology, Abbott



with either low viral loads or limited sample volume necessitated NGS sensitivity improvements. Target capture methods (xGen) were first applied to HPgV-2 and then developed for HIV, hepatitis B and D, and SARS-CoV-2 viruses (*37*–*40*). HIV probes proved instrumental in amplifying reads from a precious remnant sample to determine a third full-length sequence that finally established HIV-1 subtype L (*41*). This sensitive method was exploited once again to identify a high percentage of HIV elite controllers—defined as HIV+ individuals maintaining no detectable viral load for 12 months with no antiviral treatment—in the Democratic Republic of Congo (*42*). Hybridization of probes against all human viruses represents not only a 1000× boost in sensitivity but also allows high multiplexing and rapid data analysis, as the myriad of human background reads are washed away and not sequenced (*43*). This shortcut, in parallel with mNGS for detection of nonviral and divergent viral reads, is now a pillar of APDC patient sample characterization.

In addition to NGS optimization, streamlining laboratory methods and developing new assays applied either upstream or downstream of sequencing has been another key priority. This has entailed staffing a deep bench knowledgeable and experienced in molecular [e.g., quantitative PCR (qPCR) and isothermal amplification] and serologic detection (e.g., high-volume, rapid tests). Our biosafety level (BSL)-3 facility, used to study high-risk pathogens, has proven indispensable in our response to COVID-19 (*44*, *45*). Whether working with a known pathogen or culturing newly discovered viruses, this BSL-3 lab gives us the ability to respond quickly to shifting landscapes (e.g., new SARS variants of concern) and explore virus biology to make meaningful scientific contributions (*46*). APDC is equally focused on data-analysis infrastructure.

Our bioinformatics team is building a new metagenomics and discovery pipeline (DiVir) with a web interface that harnesses the power and scalability of cloud computing. Phylogenetics expertise is another area we have invested in. From rewriting the taxonomic classification of picobirnaviruses (PBVs) to discerning the origins of SARS-CoV-2 strains circulating in Senegal, applying state-of-the-art analyses to our data is paramount (*47*, *48*).

Assembling these various elements has culminated in two recent reports highlighting these strengths. A novel PBV was discovered in a patient hospitalized in Barranquilla, Colombia with acute respiratory illness (ARI) (*49*). Development of an RT-qPCR on the *m*2000 system led to the identification and subsequent sequence characterization of 25 additional strains from the United States and Colombia. Phylogenetic analysis of the RNA-dependent RNA polymerase enzyme indicated the emergence of two lineages that clustered together as a distinct genotype: One was linked to the index case and the other to Asian strains previously associated with ARI. A "novel" orthobunyavirus was also found in an HBV+ patient in the Democratic Republic of Congo, which after 50 years in a freezer, was recently identified by the U.S. Centers for Disease Control and Prevention as Bangui virus from the Central African Republic (*50*). While related bunyaviruses have been reported to cause disease, RT-qPCR screening with a newly developed test failed to identify additional cases of infection in Africa. Phylogenetic analysis of this and other members of the *Anopheles* A serogroup reveal these are ancient viruses that infrequently spill over to humans. These and other viruses we have yet to describe are unlikely to cause the next pandemic, but we have the tools in place to make these discoveries and develop the diagnostics to detect them.

Virus discovery within the APDC network

A fundamental promise of APDC is the sharing of resources, expertise, and samples so that together we can remain vigilant while also making new discoveries. To ensure standardization of data input, we are first defining diagnostic criteria for case finding of "illnesses of unknown origin." Key to this enrollment process are universal prescreening methods (e.g., qPCR) to flag common pathogens involved in acute respiratory or febrile illnesses, to name a few conditions. Abbott is also providing protocols and resources for conducting mNGS and target enrichment, thereby building sequencing capacity in-country. In some cases, specimens are shipped to Abbott for processing, while in others, both parties sequence the same sample to ideally arrive at the same result. Transfer of panels containing positive controls ensures laboratory methods are working as expected, yet here again, this is only half the equation. Abbott has designed the DiVir pipeline with remote, cloud-based data upload and visualization for our APDC partners. This will allow the end-user to analyze their own data and to be the ones to drive their research forward. Similarly, by providing Abbott instrumentation, these sites will also be versed in their operation and ready to test or develop their own "research use only" (RUO) assays on these platforms.

Our identification of an outbreak of Oropouche fever in Colombia is a prime example of just how these collaborations can function and provide real public health value. Fever clinics at four sites in Colombia screening for dengue, Zika, chikungunya, Chagas, and other viruses designated undiagnosed cases for further NGS analysis. An acute Oropouche bunyavirus infection was detected, and the full-length sequence indicated that the virus had been circulating undetected in Colombia for years after an initial case report in 2017 (*51*). Both an RT-qPCR assay on the *m*2000 system and immunoglobulin M/immunoglobulin G (IgM/IgG) serologic assays on the ARCHITECT platform were rapidly developed to screen nearly 1,000 fever samples collected between 2019 and 2021. We observed an 11% incidence of new infections and 4.5% IgM/15.2% IgG antibody positivity. This work highlighted the importance of logistics and the need for infrastructure to be in place to enable a rapid response. Whether it's sample collection, overseas shipments, instrumentation placements, or having sequencing and qPCR/serology capabilities at Abbott and abroad, working out these details and actively building technical capacity in advance of an outbreak makes all the difference for timely reporting and preparedness.

Preparedness for a rapid, informed response

Diagnostics are unquestionably at the center of any pandemic response strategy: Every aspect of clinical management and a public health response begins with a positive or negative test. Whether tallying up new cases to determine infection rates, monitoring how long a patient is shedding infectious virus to determine quarantine lengths, or looking at induction and maintenance of antibody responses following immunization and natural exposure, a diagnostic is the starting point. Indeed, tracking new variants through sequencing

and sophisticated phylogenetics requires first identifying a positive sample. For SARS-CoV-2, Abbott correlated the PCR cycle threshold (Ct, the number of cycles required to detect the target above the background) with virus growth in culture to indicate the viral load likely required for transmission and hence the sensitivity demanded of a rapid test like BinaxNOW (*45*). Similarly, we and our partners dissected discordant molecular and rapid (Panbio) test results to establish the window in which a person is infectious (*52*). These types of studies with on-market tests lead to an appropriate interpretation of multiple data points and an informed public health response. While proof-of-principle tests and academia-led initiatives are admirable and needed, the ability of industry to rapidly deliver reliable products in mass quantities that adhere to government quality standards and specifications (e.g., U.S. Food and Drug Administration approval and European CE marking, among others) is frankly what the world expects and exactly what Abbott delivered during the COVID pandemic.

Setting up the Virus Discovery program across APDC sites is intended to serve as a "sentinel system." This research is often referred to as "looking for the needle in the haystack," but in another way, it's more like playing the lottery: You have to buy a ticket to win. Our chance of success increases with each additional site exploring more patient samples, applying the most sensitive tools available, and following a well-thought-out strategy. If a new threat emerges, our network sites function within or hand-in-hand with hospitals that have access to patients. Regardless of whether APDC members make the initial discovery or read about the next potential pandemic in the news, we are poised to act together immediately. Methods, training, instrumentation, and funding will already be in place to sequence and comprehend genetic diversity; design

"In the first year as an APDC partner, we've worked with Abbott to phylogenetically characterize over 1,000 SARS-CoV-2 sequences to understand how the Mu variant overtook Gamma, while also screening over 1,000 patients with fever of unknown origin. In addition to our work, other sites are also actively publishing research, identifying outbreaks, and developing diagnostic tools. By copublishing and sharing patient samples, staff, and technology, we are able to work together more collaboratively and efficiently, resulting in a greater mutual benefit to our members and to public health."

—Jorge Osorio, codirector of the Colombia-Wisconsin One Health Consortium and president and CEO of VaxThera

and produce diagnostic tests; and understand the epidemiology. With testing distributed across sites, RUO assays can quickly progress to scaled-up versions with diagnostic authorization. APDC, at its core, is a network of action. Abbott is not here to be a funder, but rather as an engaged, equal partner: We are making an investment *and* doing the hard work, the results of which are already yielding benefits.

References

1. C. Y. Chiu, *Curr. Opin. Microbiol.* **16**, 468–478 (2013).
2. A. L. Greninger, *Expert Rev. Mol. Diagn.* **18**, 605–615 (2018).
3. R. Schlaberg *et al.*, *Arch. Pathol. Lab. Med.* **141**, 776–786 (2017).
4. M. R. Wilson *et al.*, *N. Engl. J. Med.* **380**, 2327–2340 (2019).
5. A. L. Greninger, *Virus Res.* **244**, 218–229 (2018).
6. S. J. Anthony *et al.*, *mBio* **8**, e00373–e00317 (2017).
7. S. J. Anthony *et al.*, *mBio* **6**, e01180–e01115 (2015).
8. T. J. Divers *et al.*, *Emerg. Infect. Dis.* **24**, 303–310 (2018).
9. T. Goldstein *et al.*, *Nat. Microbiol.* **3**, 1084–1089 (2018).
10. W. T. He *et al.*, *Cell* **185**, 1117–1129.e8 (2022).
11. C. X. Li *et al.*, *Elife* **4**, e05378 (2015).
12. M. Shi *et al.*, *J. Virol.* **90**, 659–669 (2016).
13. R. C. Edgar *et al.*, *Nature* **602**, 142–147 (2022).
14. W. I. Lipkin, *Nat. Rev. Microbiol.* **11**, 133–141 (2013).
15. J. H. Epstein, S. J. Anthony, *Rev. Sci. Tech.* **36**, 499–512 (2017).
16. J. N. Simons *et al.*, *Nat. Med.* **1**, 564–569 (1995).
17. R. Schlaberg, D. J. Choe, K. R. Brown, H. M. Thaker, I. R. Singh, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16351–16356 (2009).
18. A. Urisman *et al.*, *PLOS Pathog.* **2**, e25 (2006).
19. S. C. Lo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15874–15879 (2010).
20. V. C. Lombardi *et al.*, *Science* **326**, 585–589 (2009).
21. K. Delviks-Frankenberry, O. Cingoz, J. M. Coffin, V. K. Pathak, *Curr. Opin. Virol.* **2**, 499–507 (2012).
22. B. Oakes, X. Qiu, S. Levine, J. Hackett, Jr., B. T. Huber, *Adv. Virol.* **2011**, 854540 (2011).
23. E. C. Knouf *et al.*, *J. Virol.* **83**, 7353–7356 (2009).
24. T. Paprotka *et al.*, *Science* **333**, 97–101 (2011).
25. S. N. Naccache *et al.*, *Clin. Infect. Dis.* **60**, 919–923 (2015).
26. G. Grard *et al.*, *PLOS Pathog.* **8**, e1002924 (2012).
27. A. Swei *et al.*, *PLOS One* **8**, e62083 (2013).
28. G. Yu *et al.*, *PLOS One* **7**, e49449 (2012).
29. M. R. Wilson *et al.*, *N. Engl. J. Med.* **370**, 2408–2417 (2014).
30. S. N. Naccache *et al.*, *Genome Res.* **24**, 1180–1192 (2014).
31. M. G. Berg *et al.*, *PLOS Pathog.* **11**, e1005325 (2015).
32. K. E. Coller *et al.*, *J. Clin. Microbiol.* **54**, 2023–2030 (2016).
33. M. Frankel *et al.*, *J. Virol. Methods* **241**, 34–40 (2017).
34. A. J. Kandathil *et al.*, *Ann. Intern. Med.* **167**, 1–7 (2017).
35. A. Kapoor *et al.*, *MBio* **6**, e01466–01415 (2015).
36. H. Wang *et al.*, *Clin. Infect. Dis.* **66**, 29–35 (2018).
37. M. G. Berg *et al.*, *PLOS One* **15**, e0236046 (2020).
38. K. Forberg *et al.*, *Virology* **539**, 69–79 (2020).
39. G. S. Orf *et al.*, *Front. Virol.* **1**, 765974 (2021).
40. J. Yamaguchi *et al.*, *Front. Microbiol.* **9**, 2150 (2018).
41. J. Yamaguchi *et al.*, *J. Acquir. Immune Defic. Syndr.* **83**, 319–322 (2020).
42. M. G. Berg *et al.*, *EBioMedicine* **65**, 103258 (2021).
43. T. Briese *et al.*, *MBio* **6**, e01491–01415 (2015).
44. M. G. Berg, E. Israeli, E. Quaco, G. A. Cloherty, P. M. Hemken, *Biotechniques* **71**, 370–375 (2021).
45. M. G. Berg *et al.*, *J. Clin. Virol.* **143**, 104945 (2021).
46. M. A. Rodgers *et al.*, *J. Clin. Virol.* **147**, 105080 (2022).
47. L. J. Perez, G. A. Cloherty, M. G. Berg, *Viruses* **13**, 1476 (2021).
48. L. J. Perez *et al.*, *Virus Evol.* **8**, veac025 (2022).
49. M. G. Berg *et al.*, *Viruses* **13**, 2534 (2021).
50. M. L. Kapuscinski *et al.*, *PLOS Pathog.* **17**, e1009315 (2021).
51. D. E. Gomez-Camargo *et al.*, *Emerg. Infect. Dis.* **27**, 1756–1758 (2021).
52. I. A. Corrêa *et al.*, *medRxiv* (2022), https://doi.org/10.1101/2022.03.17.22272008.

The key role of field epidemiology in preventing the next pandemic: The partnership between Abbott and Training Programs in Epidemiology and Interventions Network (TEPHINET)

Francisco Averhoff[1], Lara Pereira[2], Susan Gawel[1]

On March 11, 2020, prompted by the detection of over 118,000 cases, 4,000 deaths, and alarming levels of spread and severity in more than 100 countries and territories, the World Health Organization (WHO) declared the novel coronavirus 2019 (COVID-19) outbreak a global pandemic (*1*). After that pivotal announcement, the absence of a coordinated global preparedness plan was evident, highlighting the need to focus on pandemic preparedness for the *next* pandemic. The COVID-19 pandemic also underscored the critical need for disease surveillance as well as the importance of trained frontline public health professionals, such as field epidemiologists and clinicians, who can recognize unusual medical events or clustering of medical conditions or syndromes. Further, the global spread and impact of the SARS-CoV-2 virus demonstrated that coordination and collaboration among public health networks regionally and worldwide are essential to ensure a rapid and robust response to infectious disease threats. Abbott, and the Abbott Pandemic Defense Coalition (APDC), recognized the need to partner with and support epidemiologic programs as key in early detection of and response to pandemic threats.

In the 1980s, several countries with funding and technical support from the U.S. Centers for Disease Control and Prevention (CDC) and other partners developed field-based training programs in applied epidemiology in response to the need to strengthen their local and national public health capacity and infrastructure. The benefits resulting from these

[1]Abbott, Abbott Park, IL
[2]Task Force for Global Health, Decatur, GA

PHOTO: NZOGANG DJOUOKA COLETTE

Supervision of a team during the malaria outbreak investigation in 2020 at the Minawao refugee camp, Far North Region, Cameroon.

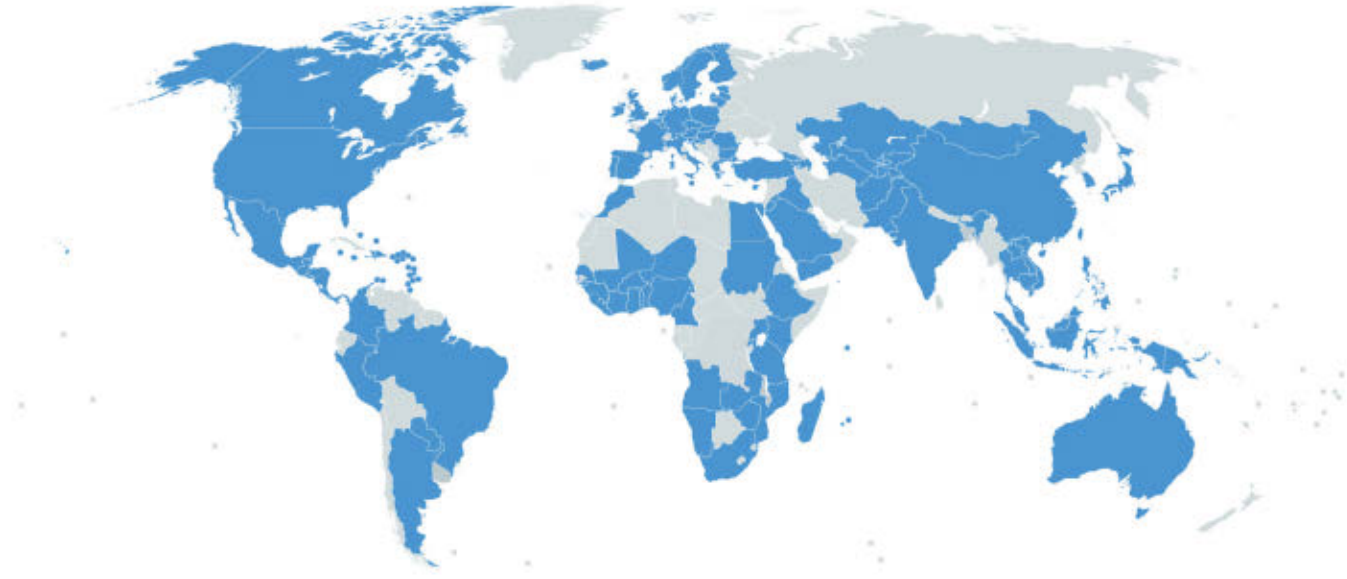


FIGURE 1. Countries with Training in Epidemiology and Public Health Interventions network (TEPHINET)-affiliated Field Epidemiology Training Programs (FETPs) (*n*=75), September 2021.

programs working together prompted the organization of a global network and, with support from WHO, CDC, and Fondation Mérieux, this network was formalized in 1997 as the Training Programs in Epidemiology and Public Health Interventions Network (TEPHINET). TEPHINET is a program of The Task Force for Global Health, originally founded as the Task Force for Child Survival by Dr. Bill Foege in 1984 (*2*). TEPHINET has grown into a global network of 75 Field Epidemiology Training Programs (FETPs) in more than 100 countries (see Figure 1). TEPHINET partners with local and national government ministries, academia, nonprofits, nongovernmental organizations, and now, through this partnership with Abbott and the APDC, with the private sector.

The TEPHINET/FETP model—"training through service"—embraces a culture of shared expertise and resources, which facilitates the program's alignment with regional and in-country public health needs to ensure sustainable impact. Training of FETP Fellows is achieved through a three-tiered approach that addresses frontline (local), intermediate (state), and advanced (national and academic) training needs, mirroring the public health infrastructure in many countries. The program confers either a certificate of completion or a Master of Public Health (MPH) degree, depending on the training level and in-country university collaborations. TEPHINET supports core competency domains that encompass both infectious and noncommunicable diseases, including surveillance, outbreak investigations, epidemiologic methods, scientific communication, and mentorship. Key to TEPHINET's mission is the development and dissemination of learning resources for both trainers and trainees, which include curriculum, knowledge hubs for high-priority topics, an abstract database, tools for virtual and in-person instruction, and hosting of emergency response trainings. Additionally, TEPHINET accreditation of FETPs ensures standardization of quality training and increased recognition of the FETPs' value in addressing country-specific health priorities. TEPHINET has trained more than 19,000 public health professionals across the world. FETP trainees and alumni serve as "boots on the ground" in the fight for better health and have led, developed, and/or implemented disease surveillance systems, outbreak investigations, and mitigation of acute health events, including playing vital roles in the COVID-19 response in many countries (*3*, *4*).

While the timing, location, and severity of the next pandemic cannot be predicted, it will be frontline field epidemiologists with expertise in surveillance and outbreak investigations who will serve as the eyes and ears to identify and define the clinical, laboratory, and epidemiologic characteristics of the next emerging pathogen of pandemic potential, and, critically, alert and inform the world of the potential threat. Epidemiologists working seamlessly with microbiologists and virologists will be the backbone of early detec-

tion and control of future pandemic threats. This need gave impetus in 2021 to APDC and TEPHINET entering a first-of-its-kind, industry–public health partnership to support and strengthen epidemiologic capacity in pandemic preparedness in low- and middle-income countries. The APDC-FETP Fellowship, launched in 2021, provides funding, mentorship, and networking opportunities to TEPHINET-member FETP trainees and graduates to conduct innovative projects in pathogen discovery, surveillance, and other priority public health focus areas in collaboration with APDC partners in their home country (*5*). During the first year of the partnership in 2021, of 72 competitive applications, eight projects were launched in seven countries in South America, Africa, South Asia, and the Caucasus. (see Table 1). These projects are primarily focused on developing the infrastructure needed to establish systems for early detection of pandemic threats and/or studying infectious diseases of public health importance in the host country. Fellows are awarded a small grant to fund project implementation. Each fellow is required to identify a local mentor and institution(s) that will host and/or support implementation of their proposed project while also ensuring that their government ministries are aware of and amenable to the institution's participation or collaboration in the project. In addition, an international subject matter expert in either epidemiology or infectious diseases is recruited to complement and strengthen the fellow's mentorship experience. As of this writing, the projects are ongoing, with a completion date in late 2022.

Since FETP programs are often associated with the country's Ministries of Health and/or National Public Health Institutes, an additional benefit of the APDC-FETP Fellowship is the bringing together of public health institutions with APDC partners in the country or region; this would provide government entities potential access to partners, the APDC, and Abbott, which have substantial capacity in pathogen discovery. These linkages may foster public–private partnerships that will be critical for timely response to future pandemic threats by strengthening collaborative laboratory and epidemiologic relationships nationally, regionally, and globally.

References

1. D. Cucinotta, M. Vanelli, *Acta Biomed.* **91**, 157–160 (2020).
2. Training Programs in Epidemiology and Interventions Network (TEPHINET) (website); https://www.tephinet.org.
3. R. Martin, I. S. Fall, *Int. J. Infect. Dis.* **110**, S3–S5 (2021).
4. A. Hilmers, *Int. J. Infect. Dis.* **110**, S1–S2 (2021).
5. F. Averhoff *et al.*, *Int. J. Infect. Dis.* **117**, S356–S360 (2022).

Country	Title
Colombia	Surveillance of respiratory agents in acute respiratory infections in Antioquia, Colombia
Georgia	(a) Association of hepatocellular carcinoma (HCC) and hepatitis C virus (HCV) in Georgia; (b) establishment of pathogen discovery surveillance in Georgia
Brazil	Respiratory and arbovirus surveillance in slum, Brasília
Tanzania	Re-establish noninfluenza respiratory surveillance, Tanzania
Bangladesh	Establishing laboratory capacity for West Nile virus (WNV) in acute febrile illness (AFI) and acute encephalitis syndrome surveillance systems, Bangladesh
Nigeria	Piloting the establishment of a dengue fever surveillance system in Bayelsa State, Nigeria
Nigeria	Meningococcal meningitis: Genomic sequencing of meningococcus serogroup X in the meningitis outbreak, Northern Nigeria
Uganda	Establishing sentinel surveillance for blackwater fever (BWF) in Uganda

TABLE 1. Training in Epidemiology and Public Health Interventions Network (TEPHINET)-Abbott/Abbott Pandemic Defense Coalition (APDC) Fellowship Awards, 2021–2022.

IMAGE: © TORIA/SHUTTERSTOCK.COM

The case for networking the networks

Gavin Cloherty[1,2], Nicaise Ndembi[3,4], Ian Lipkin[5], Sandra Chavez[6]

Detection and epidemiological characterization of infectious disease outbreaks is key for early identification and response to potential pandemic threats. The International Health Regulations (IHR) of 2005 committed all member states of the United Nations to developing the capacity to detect infectious diseases and to report those diseases in a transparent manner. Fifteen years later, the goals of the IHR had not been achieved.

The rapid, global spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2020 highlighted the gaps in the early-warning and global-response systems that were in place prior to the pandemic. One gap in the world's ability to rapidly respond was a lack of coordination on critical aspects of infectious disease surveillance and outbreak response. The *100 Days Mission to Respond to Future Pandemic Threats* (*1*)—a goal established by the Group of Seven (G7) to harness public–private collaboration and scientific innovation to deploy diagnostics, therapeutics, and vaccines within the first 100 days of a pandemic's identification—recognized that we have seen tremendous innovations from partnerships between academia, industry, international organizations, philanthropy, and governments responding to the SARS-CoV-2 pandemic. However, industry research and collaboration were absent from active participation in prepandemic public health initiatives. To avoid repeating history, we must learn from current and previous pandemics.

The world is currently making significant investments in pandemic preparedness. In recent months, numerous pandemic preparedness initiatives have been launched around the world; some overlap in scope but vary by geography or specific area of focus, while others are tackling different yet related aspects of pandemic preparedness. These initiatives, if operating independently, share the same fundamental lack of integration that was the Achilles heel of previous efforts, including duplication of effort, wasted resources, a lack of visibility into emerging threats, and fragmented response

[1]Abbott, Abbott Park, IL, USA
[2]Abbott Pandemic Defense Coalition, USA
[3]Africa CDC, African Union Commission, Addis Ababa, Ethiopia
[4]Institute of Human Virology (IHV), University of Maryland School of Medicine, Baltimore, MD, USA
[5]Columbia University, New York, NY, USA
[6]Foundation for Influenza Epidemiology, Fondation de France, Paris, France

affecting the development of tools needed to respond quickly. If we can find a way to connect the different programs, including initiatives led by industry, the world will recognize the full potential of this historic moment.

Programs and frameworks for pandemic preparedness

In the United States, the Pandemic Innovation Task Force within the White House Office of Science and Technology Policy (OSTP) was established to develop new countermeasures against possible future COVID-19 variants and other pandemic threats. The task force is charged with addressing all aspects of pandemic preparedness, including vaccination, therapeutics, diagnostics, and other interventions to be used in response to emergent variants or other public health threats. The need for such programs dedicated to the preparation for and response to future pandemics is critical if we are to avoid the devastating impact that responses to emerging threats have had on treatment of already existing epidemics and pandemics. This includes diseases such as HIV and syphilis that have increased sharply due in no small part to the tremendous strain the SARS-CoV-2 pandemic has placed on a stretched and neglected public health system.

Several public health programs have already been established, such as the National Institute of Allergy and Infectious Disease Centers for Research in Emerging Infectious Diseases (CREID), part of the U.S. National Institutes of Health (*2*). CREID plans to develop a framework and the infrastructure necessary to respond quickly and effectively to future outbreaks by building a multidisciplinary network of teams of investigators that will conduct pathogen/host surveillance and study pathogen transmission, pathogenesis, and immunologic responses in the host. This program is intended to help develop capacity for emerging infectious disease research in high-risk and underserved regions around the world.

The U.S. Centers for Disease Control and Prevention (CDC) (*3*) has been at the forefront of pandemic preparedness and response globally for decades. The CDC's focus was originally on pandemic influenza. Beginning early in the millennium following the 1997 emergence of avian influenza (H1N1) in Hong Kong, CDC, while maintaining vigilance for pandemic influenza, has been a leader in responding to the HIV pandemic through the President's Emergency Plan for AIDS Relief (PEPFAR) (*4*), and to the Ebola outbreaks in Africa and the Zika outbreaks in South America, among others. CDC has pathogen-specific expertise in epidemiology, medicine, and diagnostics. It also provides leadership in systems for detection and response to emerging threats through the establishment of the Center for Preparedness and Response and the Division of Global Health Protection; these programs address not only infectious disease threats but also natural and manmade disasters. Virtually all of CDC's infectious disease and preparedness programs were mobilized in response to the emergence of SAR-CoV-2 in 2019.

In a similar vein, an academic-led initiative out of Columbia University, the Global Alliance for Preventing Pandemics (GAPP) (*5*) builds on an infrastructure developed over decades of research and service. GAPP integrates the resources and insights of well-established subject matter experts as well as academic and public health institutions to address the urgent challenge of emerging infectious diseases. This alliance has partners in regions where pandemic threats emerge and re-emerge due to anthropogenic factors such as climate change, deforestation, displacement of wildlife, bushmeat hunting and consumption, wet markets, urbanization and migration, and incursion into wildlife habitats. It aims to develop a model that realizes and extends the goals of the IHR by providing inexpensive, rapid tools for diagnosis, discovery, and surveillance of infectious diseases. GAPP identifies and prioritizes infectious agents based on pandemic risk and shares data in a manner that provides the foundation for producing drugs and vaccines to prevent and treat infection.

Outside the United States, as part of its vision to integrate pathogen genomics into public health surveillance and disease-outbreak response, the Africa Centres for Disease Control and Prevention (Africa CDC) (*6*), through the Africa Pathogen Genomics Initiative (Africa PGI), is collaborating with partners to strengthen laboratory networks and surveillance systems on the continent. The key components of Africa PGI include (*1*) strengthening the Africa CDC Institute of Pathogen Genomics; (*2*) building a continent-wide network of pathogen genomics and bioinformatics competency; (*3*) establishing a next-generation sequencing (NGS) Academy; (*4*) creating a data architecture that will include interoperable software and a cloud-based data repository; and (*5*) developing high-impact use cases for pathogen genomics. Africa PGI has generated over 90,000 sequences from 53 African Union member states (except Eritrea and the Saharawi Republic). Thirty-six member states have referred 29,000 specimens to Africa PGI since January 2021. A notable success of this program was the identification and characterization of the

SARS-CoV-2 Omicron variant by South African scientists at the Center for Epidemic Response and Innovation, a member of Africa PGI and the Abbott Pandemic Defense Coalition (APDC).

An example of public–private partnership is the French-based Global Influenza Hospital Surveillance Network (GIHSN) (*7*). GIHSN is a platform that collects standardized data from hospitalized patients with severe acute respiratory illnesses across 22 countries, including low-, middle-, and high-income areas from both hemispheres. Clinical and epidemiological data are integrated with virological data, including whole-genomic sequencing information, which are shared with public health authorities at the local, regional, and global level. The network relies on experienced researchers and virologists from various countries and has an independent scientific committee as an advisory board. Part of the funding comes from industry, including Sanofi Vaccines, Seqirus, Illumina, and Abbott. GIHSN can inform policy decisions and serve as a network for improved respiratory virus surveillance capacity building and response. In addition to its important role in influenza surveillance, this network is positioned to have a major impact on noninfluenza respiratory pathogen surveillance. By leveraging this existing network and infrastructure, noninfluenza severe acute respiratory illness could be investigated to identify novel pathogens that may have pandemic potential.

Building and sharing disease data intelligence

Apart from the initiatives and programs outlined above, others seek to leverage the vast amount of data being generated around the world every day to help predict and model potential outbreaks with pandemic potential. This includes CDC, which recently announced the opening of a new center designed to advance the use of forecasting and outbreak analytics in public health decision making. The Center for Forecasting and Outbreak Analytics was launched in April 2022 to bring together next-generation public health data, expert disease modelers, public health emergency responders, and high-quality communications to meet the needs of decision makers. The new center will accelerate access to and use of data for public health decision makers who need information to mitigate the effects of disease threats, such as social and economic disruption.

Globally, the World Health Organization (WHO) has launched the Hub for Pandemic and Epidemic Intelligence (*8*). This hub is part of WHO's Health Emergencies Programme and will be a new global collaboration of countries and partners, driving innovation to increase the availability of key data, develop state-of-the-art analytic tools and predictive models for risk analysis, and link communities of practice around the world. Critically, it will support the work of public health experts and policymakers in all countries with the tools needed to forecast, detect, and assess epidemic and pandemic risks so they can make rapid decisions to prevent and respond to impending public health emergencies. The Global Outbreak Alert and Response Network (GOARN) (*9*) is another WHO-sanctioned network comprising technical and public health institutions, laboratories, nongovernmental organizations, and other organizations that work to detect, investigate, and respond to outbreaks, and report on potential pandemic threats.

The United Kingdom is working with the WHO to launch a "global pandemic radar" to identify emerging COVID-19 variants and track new diseases around the world. The pathogen surveillance network is intended to save lives and protect health systems by spotting diseases before they cause future pandemics.

The Rockefeller Foundation launched the Pandemic Prevention Institute (PPI) (*10*), a collaborative organization with a global network of partners committed to building data sets and analytics needed to detect, mitigate, and prevent pandemics. This institute plans to work with others, such as the WHO's Hub for Pandemic and Epidemic Intelligence and the UK's Global Pandemic Radar. PPI is also accelerating work to develop analytic tools and algorithms that detect early-warning signals wherever they occur; their goal is to design technology that sees and shares the signs of potential outbreaks and supports sharing of those signals.

Integrating industry into global public health Initiatives

One element that is often neglected is the role industry could play if engaged in active collaboration—beyond the traditional funder/donor relationship—with public health and academic initiatives. One example of this potential is the APDC. Launched in early 2021, APDC is a natural extension of Abbott's work in diagnostics, virus discovery, and virus surveillance over the past three decades.

In 1994, the Abbott Global Surveillance Program was launched and continues to be active today. This program is dedicated to tracking the evolution of HIV, viral hepatitis, SARS-CoV-2, and other pathogens around the world.

In 2009, Abbott partnered with the University of California, San Francisco, to create the first-of-its-kind, nonprofit viral diagnostics center, known as the Viral Diagnostics and Discovery Center (VDDC), to identify unknown viruses that cause acute and chronic human illnesses, including outbreaks and rare and unexplained diseases. Building on this experience and expertise, APDC is another first-of-its-kind, industry-led coalition with global, multisector scientific and public health partnerships, whose primary objective is the early detection and mitigation of infectious disease threats of pandemic potential. As of this writing, the network has partners on five continents, including academic institutions, and governmental and nongovernmental organizations. One insight that became clear during the SARS-CoV-2 pandemic is that, for diagnostics, one size does not fit all; we need a variety of tools to effectively protect public health, from simple, decentralized or at-home rapid tests, to highly sensitive, complex, and automated laboratory-based tests. No one knows what type of diagnostic tests will be needed to prevent the next pandemic. A novel element of APDC is the capacity for early development and rapid deployment of quality diagnostics. Additionally, centralized and decentralized rapid tests targeting newly identified pathogens of pandemic potential can be quickly and seamlessly produced at scale to respond to an emerging threat, at least until vaccines and therapeutics are available as the best defense.

Combining industry resources with traditional epidemiological disease surveillance tools, pathogen monitoring in animals, genomic sequencing, and artificial intelligence is vital for the creation of a global disease-warning system capable of speeding up identification of the next pandemic threat and alerting scientists and developers to develop an effective response.

Optimizing and connecting global efforts

How can we optimize the synergies of these different initiatives that share a common vision and purpose into an efficient, effective collaboration to prevent the next pandemic?

At the minimum, there needs to be open, regular communication between groups involved in pandemic preparedness. Sharing of information in the traditional way, through scientific literature or congresses, is not fast or nimble enough to meet this need. Preprint servers helped fill the gap during the pandemic; however, as submissions are neither peer reviewed nor curated, these servers provided both information and misinformation. We can do better.

Although many networks already share partner sites, each network should have a representative sitting on the steering committee of the other's network, so there is visibility not only of what is being done but also what is being planned. Information shared should include geographies covered by networks, including the sites involved and infrastructure and expertise in place as well as the specific scope and focus of the work being conducted. In this context, scientific data such as interesting cases or outbreaks—ideally including sequence, clinical, and metadata—could also be rapidly disseminated throughout a "network of networks." The rapid sharing of this type of information, especially together with samples, would facilitate rapid production of diagnostic tests that could be distributed throughout the greater network of networks to expedite further investigation and/or response.

Governments, philanthropic institutes, and industry are making record investments in public health initiatives aimed at pandemic preparedness. We must nonetheless understand that these resources are finite and subject to "panic and neglect" cycles of funding. A coordinated effort that leverages these assets would reduce waste and increase the impact of the investment. To demonstrate to funding agencies that these initiatives are sustainable and worthy of long-term support, we should also emphasize their public health benefits *between* pandemics. We can keep these networks dynamic by extending surveillance to antimicrobial resistance and food security and by investing in interventions that reduce the morbidity and mortality of infections that result in chronic and acute diseases.

References

1. Pandemic Preparedness Partnership, *100 Days Mission to Respond to Future Pandemic Threats* (2021); available at https://www.gov.uk/government/publications/100-days-mission-to-respond-to-future-pandemic-threats.
2. Centers for Research in Emerging Infectious Diseases (website); https://creid-network.org.
3. U.S. Centers for Disease Control and Prevention (website); https://www.cdc.gov.
4. The United States President's Emergency Plan for AIDS Relief (website); https://www.state.gov/pepfar.
5. Global Alliance for Preventing Pandemics (website); https://www.publichealth.columbia.edu/global-alliance-preventing-pandemics-gapp.
6. Africa Centers for Disease Control and Prevention (website); https://africacdc.org.
7. Global Influenza Hospital Surveillance Network (website); https://www.gihsn.org.
8. World Health Organization Hub for Pandemic and Epidemic Intelligence (website); https://www.who.int/initiatives/who-hub-for-pandemic-and-epidemic-intelligence.
9. Global Outbreak Alert and Response Network (website); https://extranet.who.int/goarn.
10. Pandemic Prevention Institute (website); https://www.rockefellerfoundation.org/pandemicpreventioninstitute.

Building genomic sequencing capacity in Africa to respond to the SARS-CoV-2 pandemic

Tulio de Oliveira[1,2,3,4]*, Eduan Wilkinson[1], Cheryl Baxter[1,3], Houriiyah Tegally[1], Jennifer Giandhari[2], Yeshnee Naidoo[1], Sureshnee Pillay[2]

Genomics surveillance aims to transform public health interventions by monitoring genetic changes that impact pathogenicity, diagnostics, therapeutics, and vaccines. Monitoring the genetic changes in SARS-CoV-2 has played an important role in shaping the scientific response to the pandemic and allowed the identification of several variants of interest (VOI) and five variants of concern (VOC) to date. Although Africa accounts for only about 2.5% and 4.1% of the world's reported COVID-19 cases and deaths, respectively (*1*, *2*), two of the VOC were identified by scientists from South Africa. Here, we reflect on some of the investments and capacity development initiatives that have resulted in an exponential growth in genomic sequencing capabilities across the continent over the past 2 years.

Early in the SARS-CoV-2 pandemic, genomic surveillance was available in just a few African countries with only 5,245 SARS-CoV-2 genome sequences being made publicly available in 2020 (*3*). In 2020 and 2021, significant investments in equipment and training were made to extend the geographic coverage of sequencing within many laboratories in Africa, thus increasing surveillance capacity on the continent. These investments resulted in an exponential increase in the number of SARS-CoV-2 genome sequences produced (Figure 1). Interestingly, it took 375 days to produce the first 10,000 SARS-CoV-2 genomes, 87 days to produce the next 10,000, and just 24 days to produce the most recent 10,000 genomes. To date, almost 100,000 SARS-CoV-2 genome sequences from Africa have been shared, and 54 African countries are now contributing to SARS-CoV-2 genome sequencing.

In 2020, the World Health Organization (WHO) and the Africa Centres for Disease Control and Prevention (Africa CDC)

[1]Centre for Epidemic Response and Innovation (CERI), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa
[2]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
[3]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa
[4]Department of Global Health, University of Washington, Seattle, Washington, USA
*Corresponding author: tulio@sun.ac.za

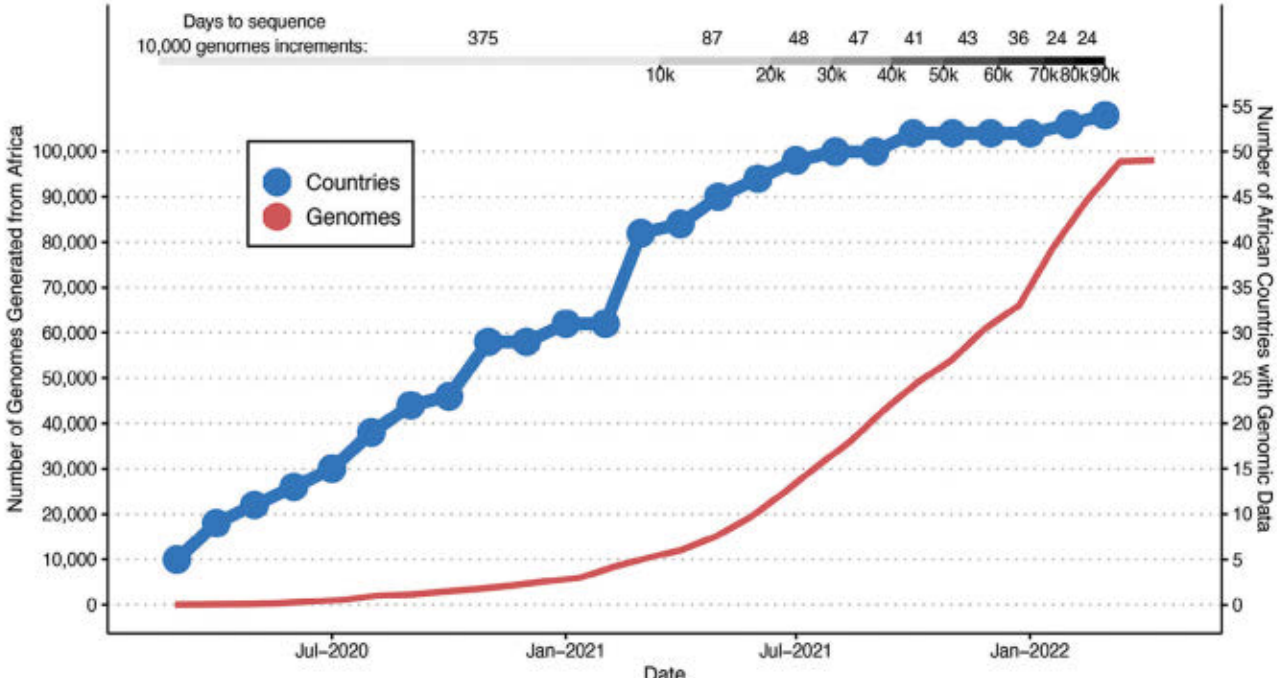


FIGURE 1. Increase in the number of SARS-CoV-2 sequences (solid red line) and African countries (blue circles) contributing genomic sequences to the Global Initiative on Sharing All Influenza Data (GISAID) between January 2020 and March 2022. Note that countries contributing to genomic sequences include 52 African countries and two overseas territories (Reunion and Mayotte).

FIGURE 2. Hands-on genomics sequencing training at the Centre for Epidemic Response and Innovation in Stellenbosch, South Africa.

launched the Pathogen Genomic Initiative (PGI), a network of laboratories established to reinforce SARS-CoV-2 genome sequencing in Africa. Through this initiative and as part of a training program funded by the Rockefeller Foundation and Africa CDC, the Centre for Epidemic Response and Innovation (CERI) and KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP) laboratories in South Africa have provided hands-on genome sequencing and analysis training to 46 fellows from 21 African countries (Figure 2). Importantly, this training has helped build local sequencing and bioinformatics capacity in several of the African countries that were previously sending samples to overseas laboratories for sequencing. In collaboration with the Africa CDC, 18 webinar sessions, which reached 123 African groups, have been hosted to train African researchers in the analysis of their own sequencing data. The first pan-African paper resulting from this collaboration was published in Science in 2021 (*4*), and another publication that highlights the 100,000 genomes generated in Africa is in preparation. An important element of the training is the transparent and open sharing of knowledge. To this end, the protocols established for genomic data generation have been made publicly available on protocols.io, and all of the data analysis scripts have been shared in GitHub. Future training events are planned through the Rockefeller Foundation and the Abbott Pandemic Defense Coalition, which is dedicated to the early detection of, and rapid response to, future pandemic threats.

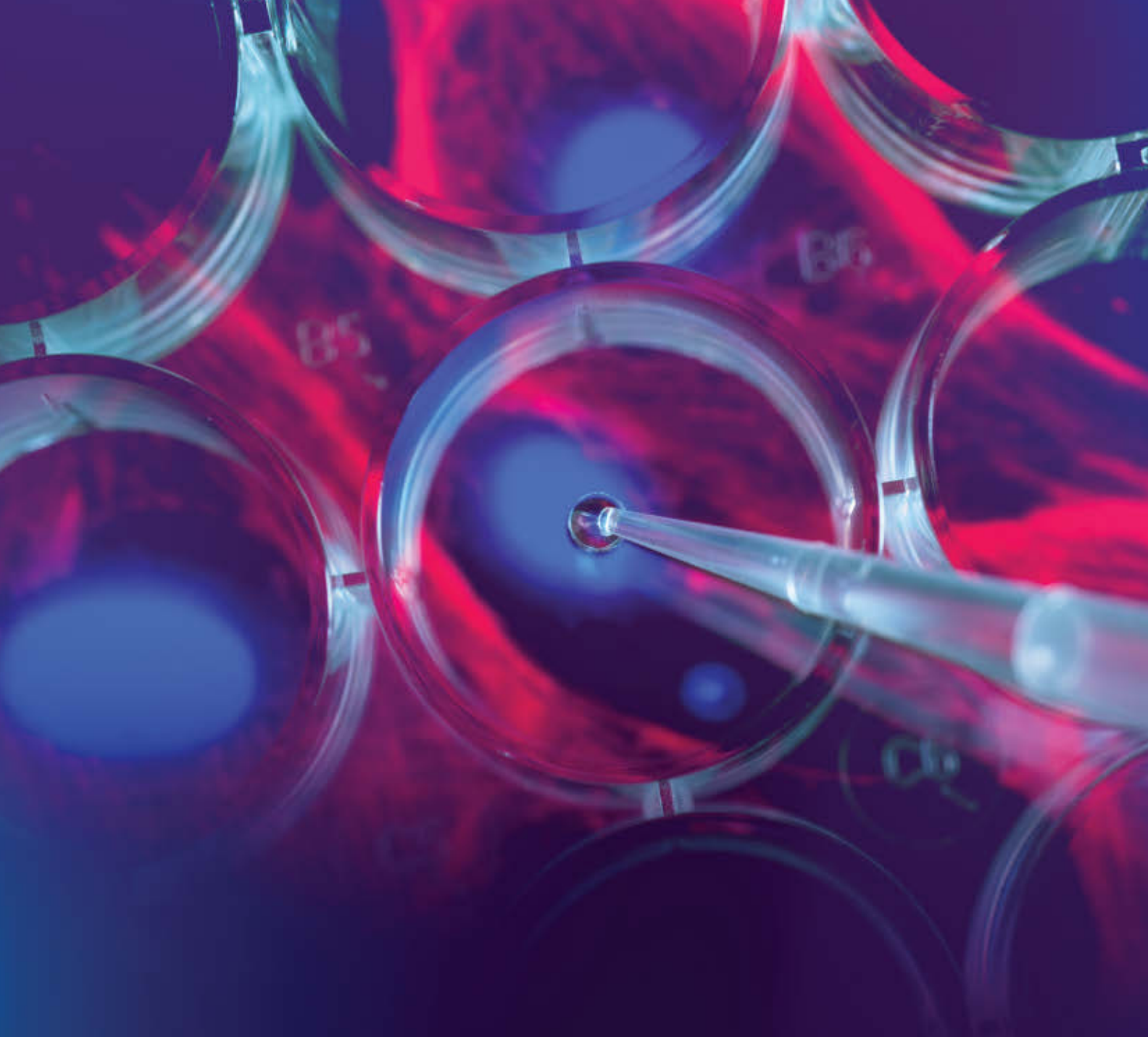
The continued generation, analysis, and sharing of virus genomes in real-time from Africa remains important to monitor the expected efficacy and sensitivity of different vaccines and nucleic acid tests across the continent. Africa's increased capacity to not only sequence viruses but also to analyze the data will surely have a multiplier effect and will be beneficial as the continent continues to grapple with ongoing and emergent pathogens in the future.

References

1. E. Dong, H. Du, L. Gardner. *Lancet Infect. Dis.* **20,** 533–534 (2020).
2. Africa CDC, "Africa CDC COVID-19 Dashboard" (2022); available at https://africacdc.org/covid-19.
3. Global Initiative on Sharing All Influenza Data (GISAID), "Tracking of Variants" (2022); available at https://www.gisaid.org/hcov19-variants.
4. E. Wilkinson *et al.*, *Science* **374**, 423–431 (2021).

HUMAN-POWERED HEALTH

The future of healthcare is now. It's human powered and human centered, driven by new technologies that are developing at an increasingly rapid pace. It's moving from reactive and retrospective to proactive and predictive. It's giving people the knowledge and the power they need to take control of their health, and live their fullest possible lives. Learn more at **abbott.com**.



THIS IS HOW WE HELP PREVENT **THE NEXT PANDEMIC**

Viruses move fast, we'll move faster. Abbott's Pandemic Defense Coalition is the first industry-led global scientific and public health partnership dedicated to the early detection of and rapid response to future pandemic threats.

Discover more at abbott.com/virushunters

Abbott